An Introduction to Bioinformatics Resources and their Practical Applications

Medha Bhagwat, PhD
Coordinator
NIH Library Bioinformatics Support Program
bhagwat@mail.nih.gov

(Really) An Introduction to *(Few)* Bioinformatics Resources and their Practical Applications

Medha Bhagwat, PhD
Coordinator
NIH Library Bioinformatics Support Program
bhagwat@mail.nih.gov

Bioinformatics

Variety of definitions

Luscombe et al. *Method Inform Med* 2001; 40:346-58

Bioinformatics is conceptualizing biology in terms of molecules
(in the sense of Physical chemistry)
and applying "informatics techniques"
(derived from disciplines such as applied
math, computer science and statistics)
to understand and organize the information associated
with these molecules, on a large scale.

Bioinformatics is a management information system for molecular biology and has many practical applications.

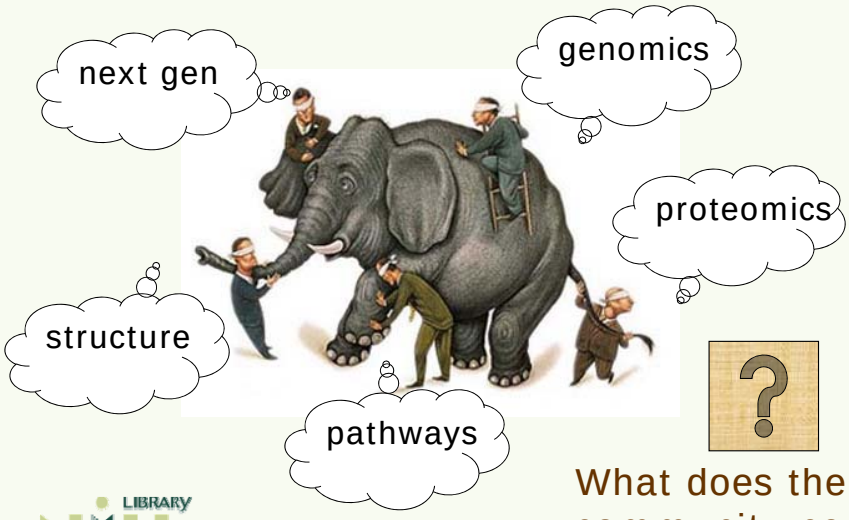
Bioinformatics

I. Organize data in databases
- access current data
- submit new data

II. Develop tools and resources to analyze data

III. Interpret data in a biologically useful manner
- global analysis of data to uncover common principles that apply across many systems

Bioinformatics



What does the NIH community need?

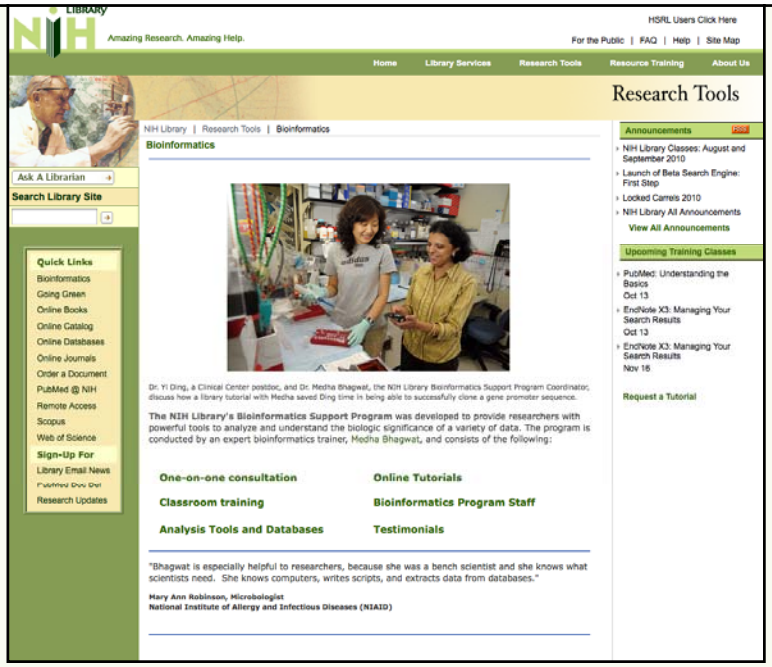
LIBRARY NIH

Identifying and Supporting NIH Researchers' Bioinformatics Needs

October 6
11AM-1PM

RSCHSUPP – 4

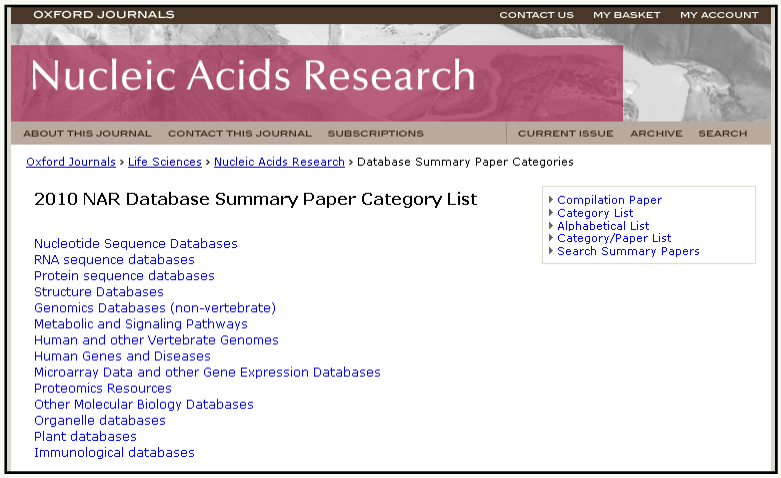
LIBRARY NIH



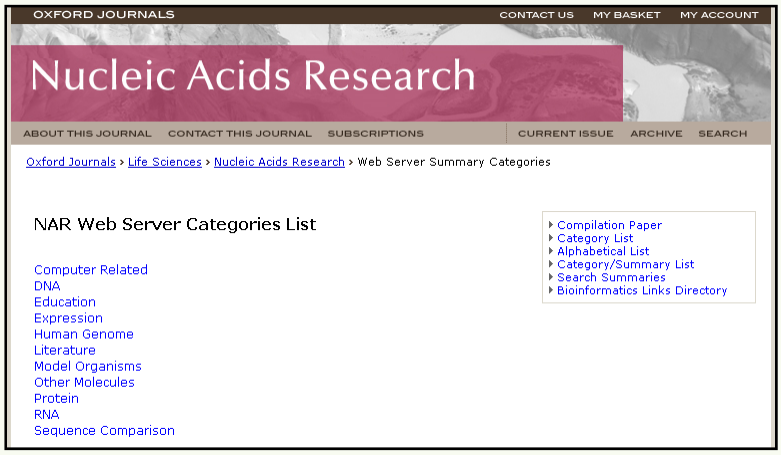
http://nihlibrary.nih.gov/bioinformatics

Outline

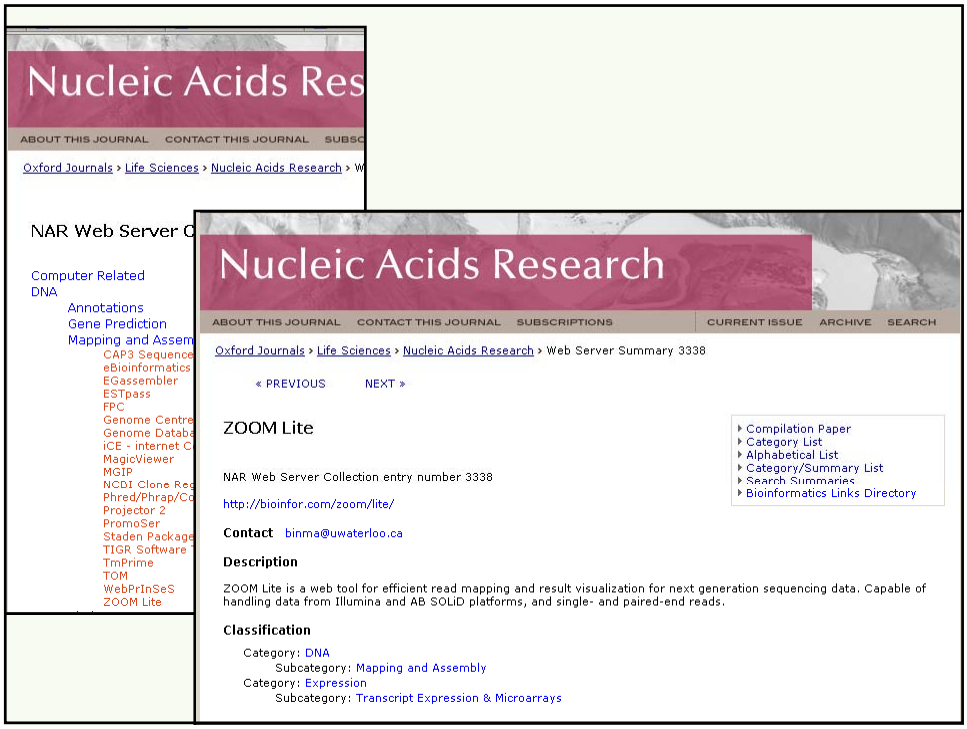
- Databases and tools (free)
- Licensed resources (from the NIH Library)
- Specific examples
- Training and additional help

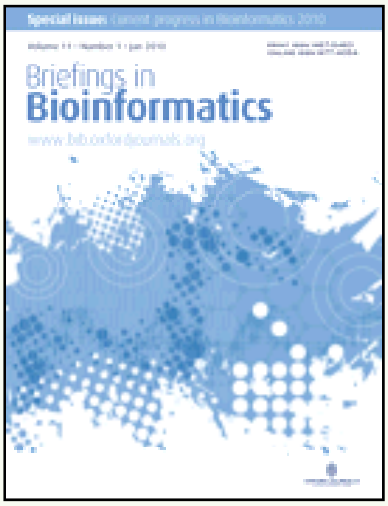


http://www.oxfordjournals.org/nar/database/c/



http://www.oxfordjournals.org/nar/webserver/c



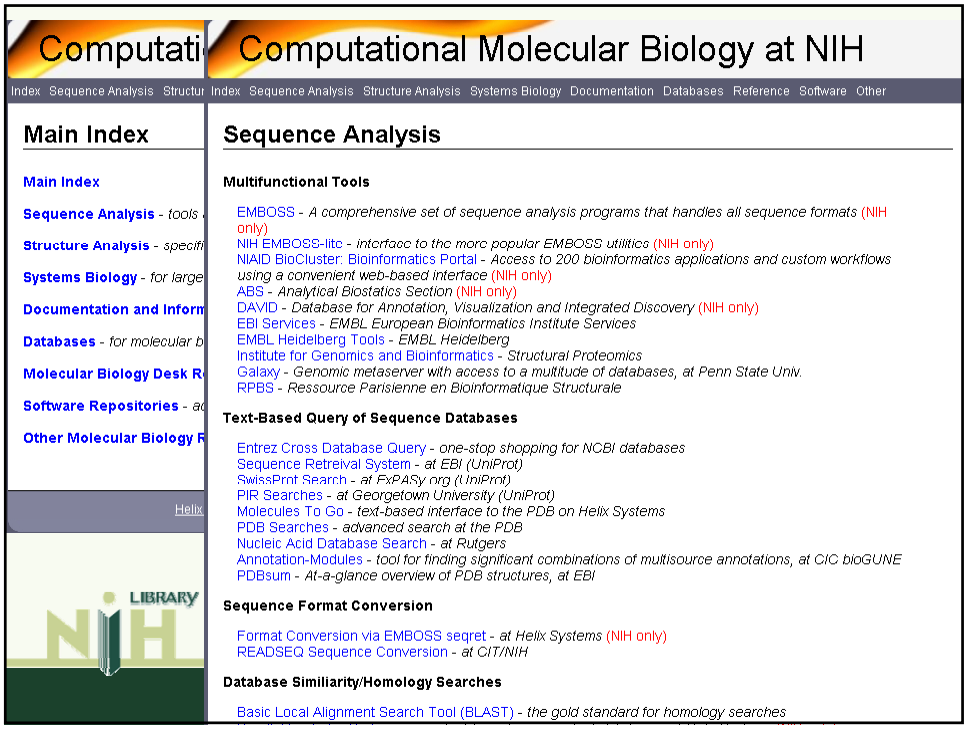


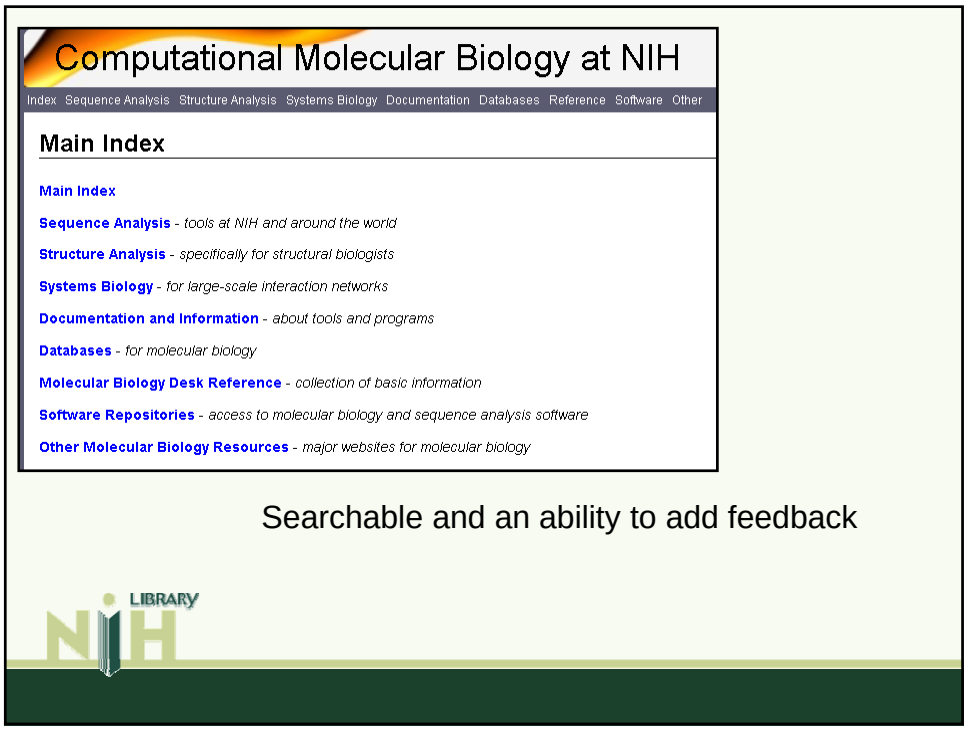
Special Issues

Current progress in Bioinformatics 2010

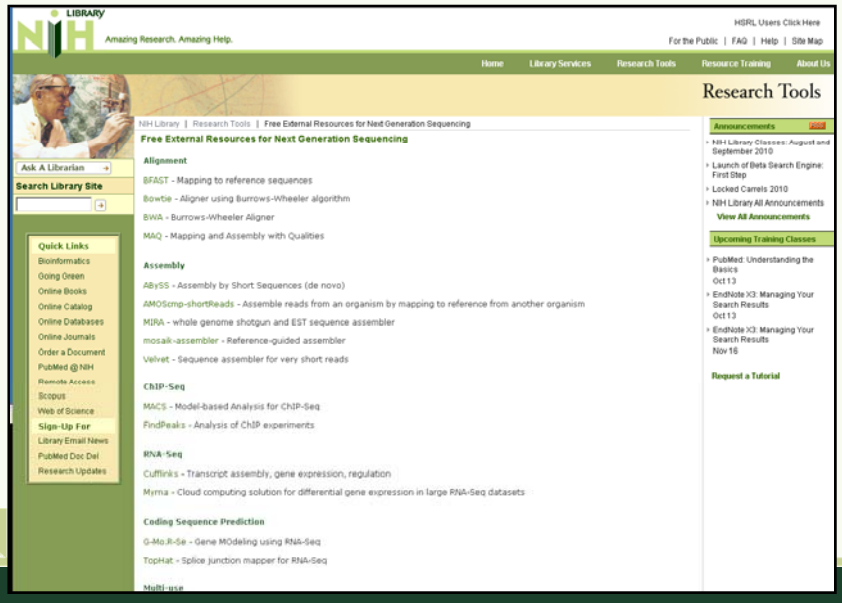
Second Generation Sequencing

LIBRARY
NIH

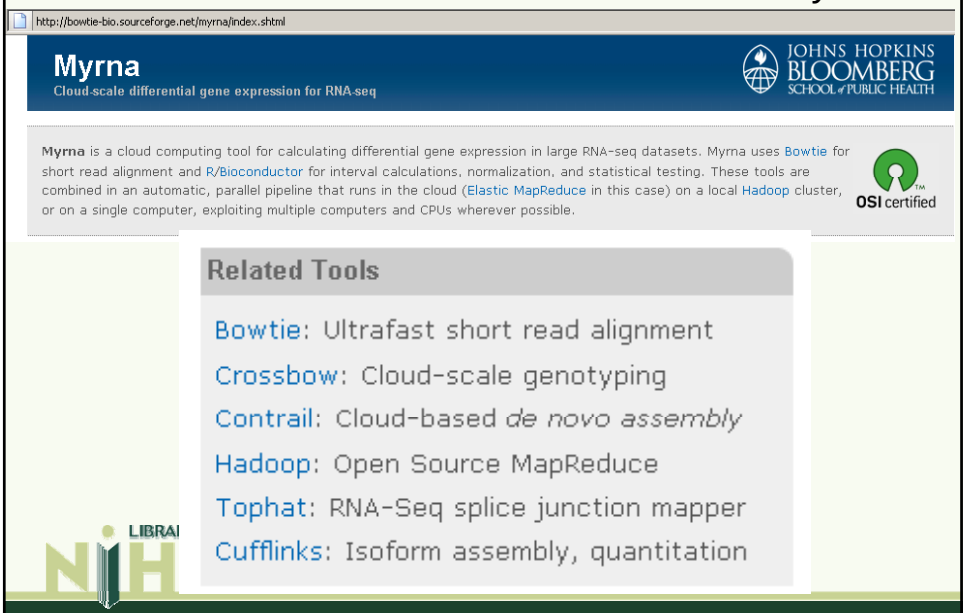

Computati
Computational Molecular Biology at NIH
Index Sequence Analysis Structur
Index Sequence Analysis Structure Analysis Systems Biology Documentation Databases Reference Software Other
Main Index
Main Index
Sequence Analysis - tools
Structure Analysis - specifi
Systems Biology - for large
Documentation and Inform
Databases - for molecular b
Molecular Biology Desk R
Software Repositories - a
Other Molecular Biology R
Helix
LIBRARY
NIH
Sequence Analysis
Multifunctional Tools
EMBOSS - A comprehensive set of sequence analysis programs that handles all sequence formats (NIH only)
NIH EMBOSS-lite - interface to the more popular EMBOSS utilities (NIH only)
NIAID BioCluster: Bioinformatics Portal - Access to 200 bioinformatics applications and custom workflows using a convenient web-based interface (NIH only)
ABS - Analytical Biostatics Section (NIH only)
DAVID - Database for Annotation, Visualization and Integrated Discovery (NIH only)
EBI Services - EMBL European Bioinformatics Institute Services
EMBL Heidelberg Tools - EMBL Heidelberg
Institute for Genomics and Bioinformatics - Structural Proteomics
Galaxy - Genomic metaserver with access to a multitude of databases, at Penn State Univ.
RPBS - Ressource Parisienne en Bioinformatique Structurale
Text-Based Query of Sequence Databases
Entrez Cross Database Query - one-stop shopping for NCBI databases
Sequence Retreival System - at EBI (UniProt)
SwissProt Search - at ExPASy.org (UniProt)
PIR Searches - at Georgetown University (UniProt)
Molecules To Go - text-based interface to the PDB on Helix Systems
PDB Searches - advanced search at the PDB
Nucleic Acid Database Search - at Rutgers
Annotation-Modules - tool for finding significant combinations of multisource annotations, at CIC bioGUNE
PDBsum - At-a-glance overview of PDB structures, at EBI
Sequence Format Conversion
Format Conversion via EMBOSS seqret - at Helix Systems (NIH only)
READSEQ Sequence Conversion - at CIT/NIH
Database Similiarity/Homology Searches
Basic Local Alignment Search Tool (BLAST) - the gold standard for homology searches


Computational Molecular Biology at NIH
Index Sequence Analysis Structure Analysis Systems Biology Documentation Databases Reference Software Other
Main Index
Main Index
Sequence Analysis - tools at NIH and around the world
Structure Analysis - specifically for structural biologists
Systems Biology - for large-scale interaction networks
Documentation and Information - about tools and programs
Databases - for molecular biology
Molecular Biology Desk Reference - collection of basic information
Software Repositories - access to molecular biology and sequence analysis software
Other Molecular Biology Resources - major websites for molecular biology
Searchable and an ability to add feedback
LIBRARY
NIH

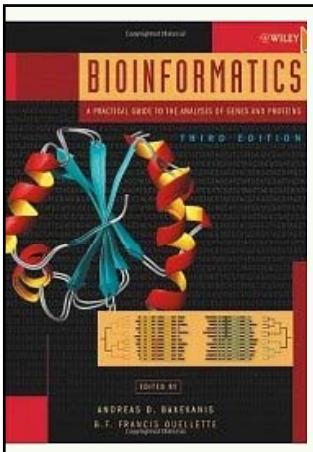
Some Free Resources for Next Gen Analysis



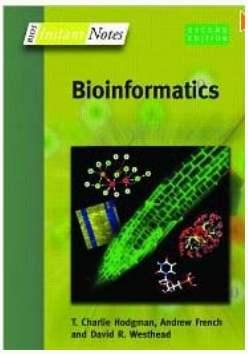
Some Free Resources for Next Gen Analysis



http://bowtie-bio.sourceforge.net/myrna



Additional help



CURRENT PROTOCOLS in Bioinformatics

LIBRARY NIH

Outline

- Databases and tools (free)
- Licensed resources (from the NIH Library)
- Training
- One-on-one consults
- Practical examples

LIBRARY NIH

Licensed Resources for NIH Staff

ArrayStar/QSeq
CLC Genomics Workbench
GeneIndexer
GeneSpring
Genomatix
**Human Genome Mutation Database Professional
Ingenuity Pathways Analysis (IPA)
Lasergene
MetaCore™ from GeneGo
Eureka!
Open Helix on-line tutorial suite
Partek Genomics Suite
ProteinLounge
SeqMan NGen

Licensed Resources (from the NIH Library)

- DNA/protein sequence analysis
- Expression data analysis
- Pathway analysis
- Next gen sequence analysis
- Promoter/SNP prediction

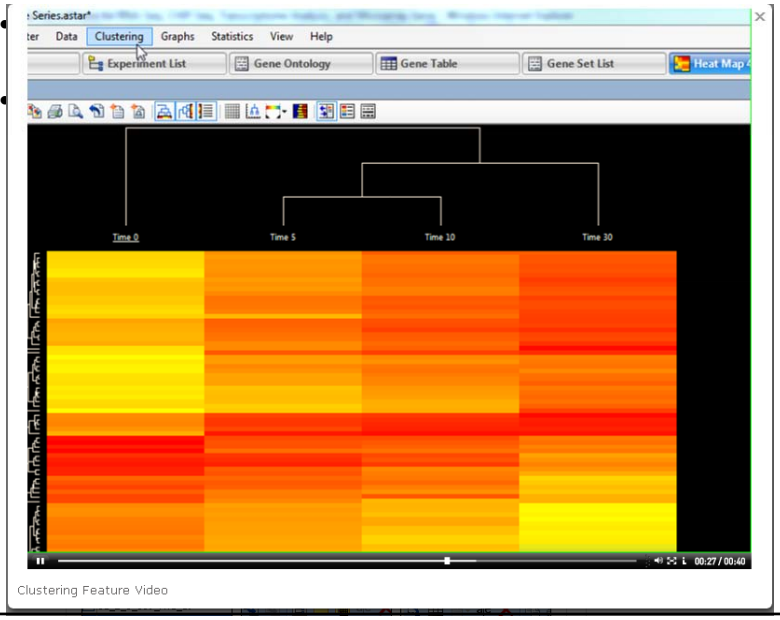
NIH LIBRARY

DNASTAR ArrayStar®

ArrayStar is a gene expression analysis and visualization software on Windows. Tools for Venn diagrams, scatter plot, heat maps and line graphs for clustering.

- Identify relationships between genes with particular biological functions
- Determine relative importance of genes in specific processes using the numerous statistical comparisons
- Download gene ontology tree structure from the Gene Ontology Consortium
- Visualize expression level changes in individual genes over the course of the experiment
- Multi-functional scatter plots generated to easily select groups of genes for analysis
- Normalization methods: RMA, PLIER, quantile normalization, or average summarization
- Cluster data using hierarchical clustering or k-means

DNASTAR ArrayStar®



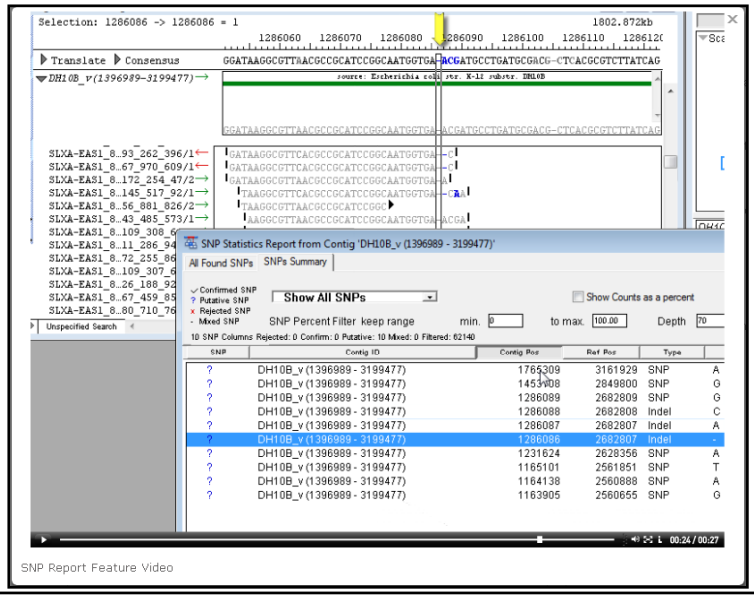
Clustering Feature Video

DNASTAR Lasergene®

DNA and protein sequence analysis on Windows and Mac
Seven applications

▪SeqBuilder	Sequence editing, annotation, automated virtual cloning, and primer design
▪SeqMan Pro	Contig assembly and analysis, including SNP discovery, coverage evaluation, and project annotation
▪MegAlign	DNA and protein sequence alignments and analysis
▪GeneQuest	Gene discovery and annotation
▪Protean	Protein structure analysis and prediction
▪PrimerSelect	Primer design
▪EditSeq	Importing and editing unusual file types

DNASTAR Lasergene®



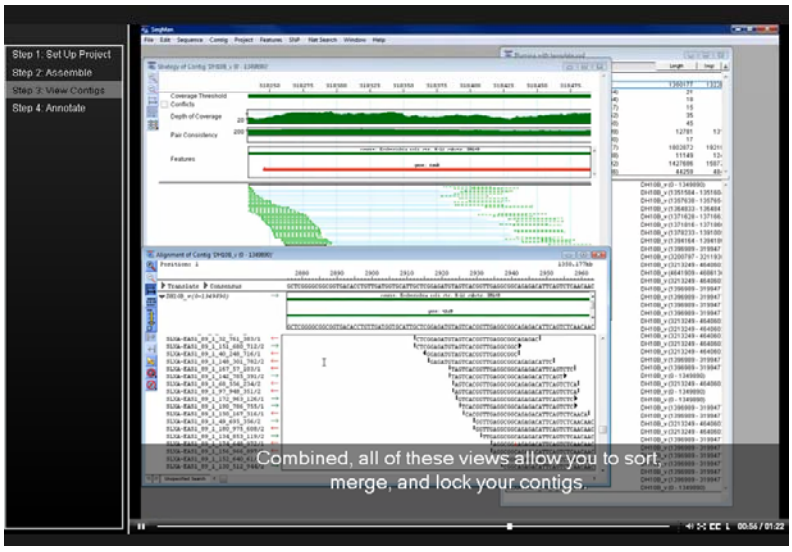
SNP Report Feature Video

DNASTAR SeqMan NGen®

SeqMan NGen on Windows and Macintosh is used for traditional, next-generation, and third-generation assemblies.

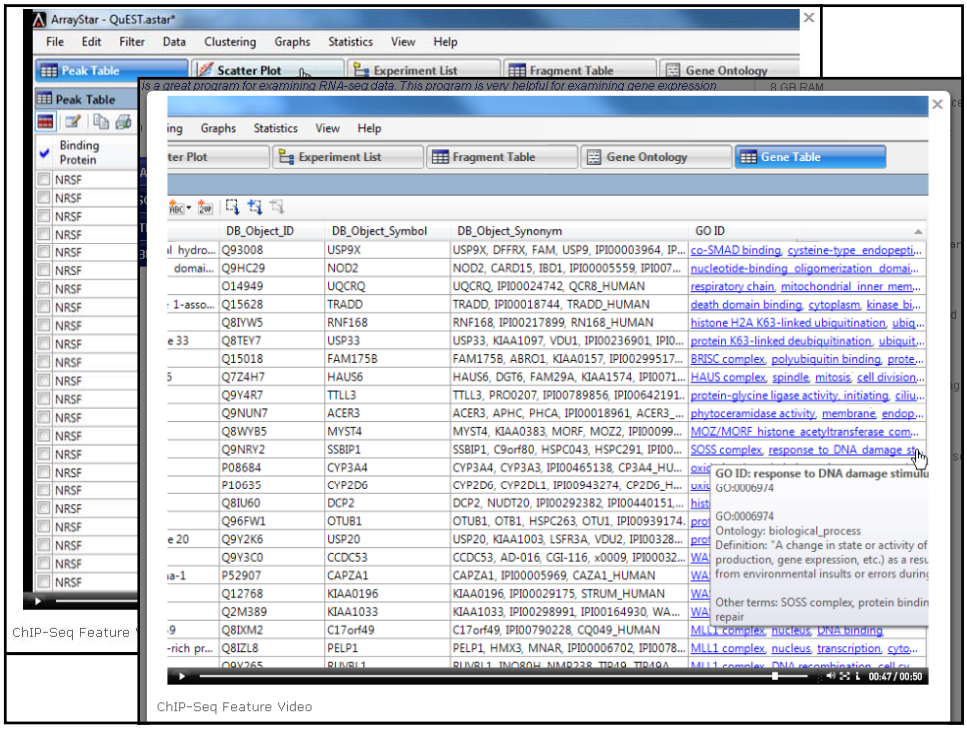
- Assembly parameter settings are determined for you based on your read technology and your specific project objective
- Integrated with Lasergene's SeqMan Pro for analyzing your project, including discovering SNPs, evaluating coverage, and annotating your consensus sequence
- Next-Gen Workflows Supported
  - *-de novo* genome, transcriptome, and metagenome fields of study
  - -Whole genome re-sequencing with gap closure, SNP analysis, and annotation
  - -Targeted re-sequencing of candidate genes or regions at high accuracy to identify low frequency SNPs and variants
  - -Assembly of ChIP-Seq and RNA-Seq reads against a reference genome

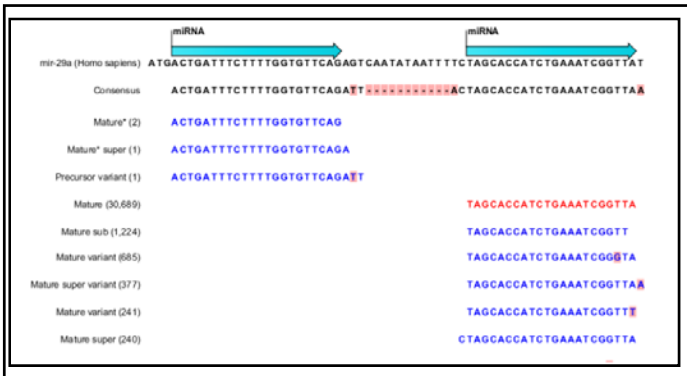
DNASTAR SeqMan NGen®



QSeq is an application for RNA-Seq, ChIP-Seq, and miRNA alignment and analysis. QSeq is fully integrated with ArrayStar, enabling you to take advantage of its powerful visualization and analytical tools, including using Gene Ontology (GO) annotations for ontology comparisons and gene characterization.

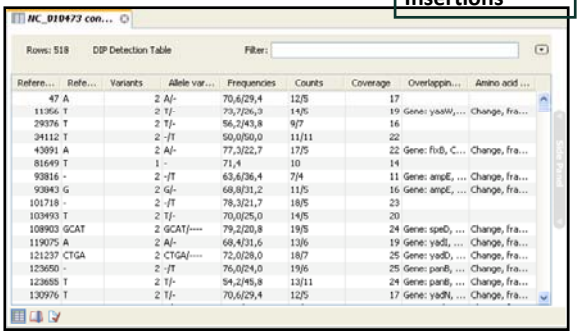
- Select gene sets and export associated reads for sequence assembly, alignment, and detailed analysis
- RNA-Seq analysis to measure gene expression for transcriptomes
- ChIP-Seq peak detection to discover binding sites of DNA-associated proteins
- Easy importation of GO annotations for ontology comparisons and gene characterization.
- High Capacity, High Speed Assembler
  - Align 1 billion or more sequences to a human genome
  - Short and long sequence reads aligned
  - Align 100 million reads in less than one hour on a common desktop machine





Alignment against the mir29a precursor sequence

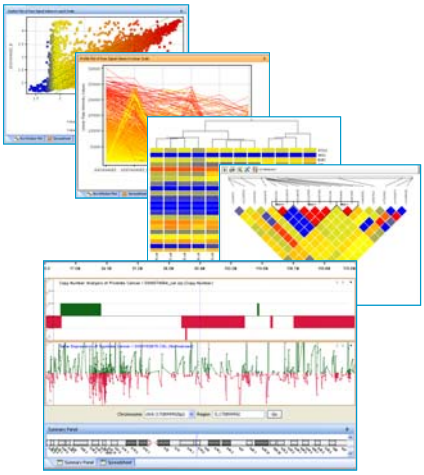
A Table of Deletions and Insertions



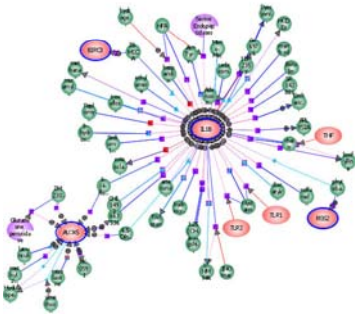
- ***de novo* Assembly** on a strong desktop computer
- **Analysis of NGS data**
- **RNA-Seq, ChIP-Seq**
- **mircoRNA analysis**
- **SNP and DIP detection**

GeneSpring GX
Platform Strengths

Powerful Statistical Tools and Sophisticated Data Visualizations



Biological Contextualization (pathway module, GSEA, GSA, GO analysis)



Open Platform
Agilent, Affymetrix, Illumina, custom
PCs, Macs, Linux

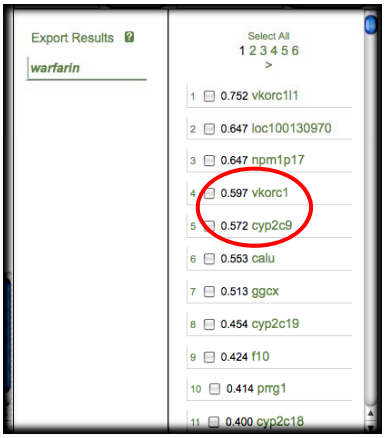
Agilent Technologies

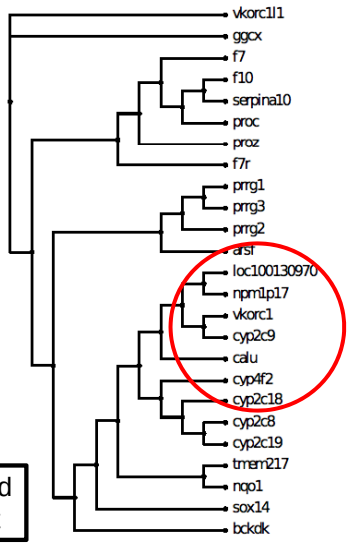
GeneIndexer

Discovery and hypotheses generation tool that compresses literature analysis from months to minutes

- Extracts implicit and explicit gene associations from over 1.5 million scientific literature abstracts
- Use free text queries (diseases, pathways, phenotypes, drugs, GO classifications, etc.) to identify and prioritize genes most relevant to a given research question.
- Builds hierarchical trees in which genes are clustered into functionally related groups.

Case Study: A Warfarin Query





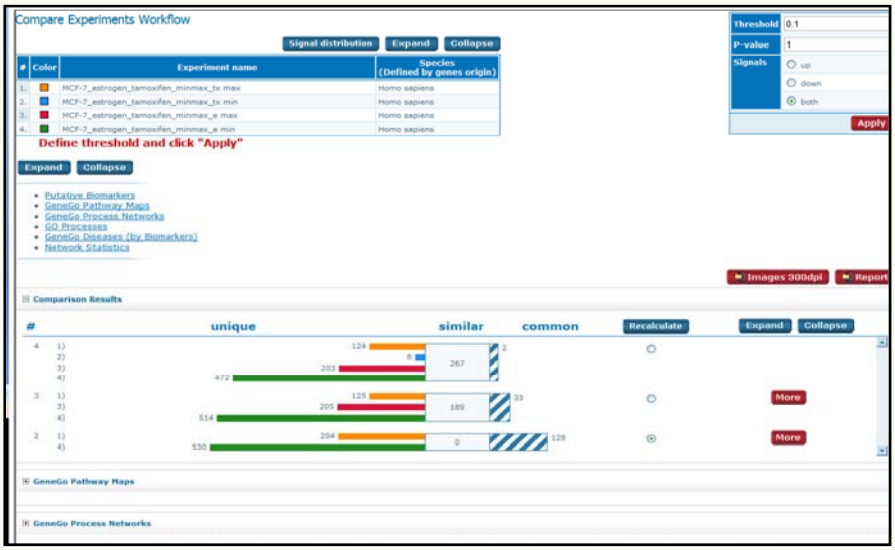
GeneIndexer quickly finds CYP2C9 and VKORC1, plus other genes of interest

MetaCore from GeneGo

- Analyze data such as microarray gene expression, SNPs, metabolic profiles, high content screening (HCS) assays
- Identify the most relevant pathways, networks and cellular processes.
- Search information about genes, proteins, compounds, pathway maps and diseases

http://www.genego.com/metacore.php

MetaCore from GeneGo



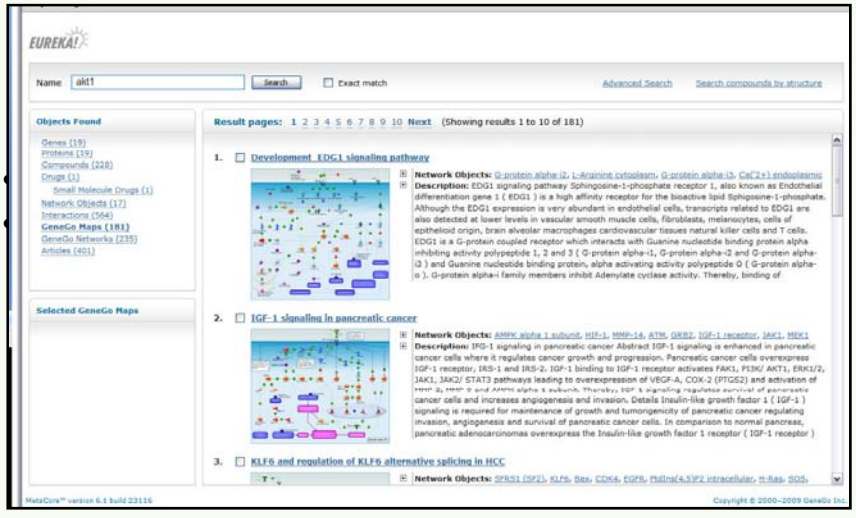
http://www.genego.com/metacore.php

Eureka!

- Search GeneGo's knowledge database
- Find pathways and enrichments for biomarkers, signaling and metabolic processes

http://www.genego.com/eureka.php

Eureka!



http://www.genego.com/eureka.php

genomatix software suite v2.0

Main menu | Logout

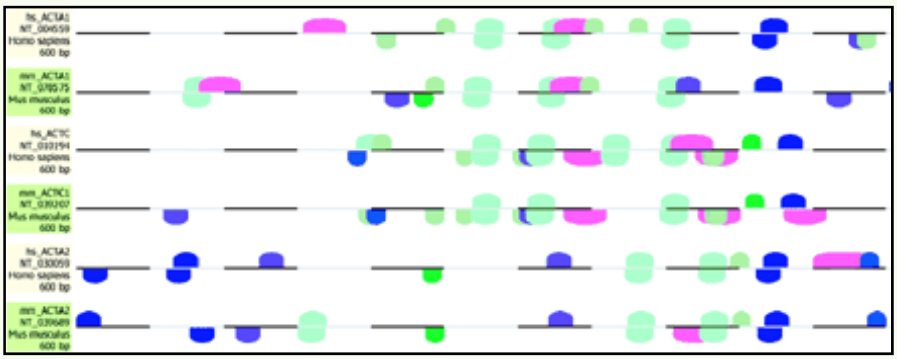
NGS Analysis | Literature & Pathways | Genomes & Data | Pattern Search & Analysis | Pattern Definition | Alignment & Mapping | Tools | Projects & Account | Help

- Analysis of genomic data, gene regulation and expression
- Generate and evaluate networks and pathways
- Perform literature searches and sequence analyses
- Visualize comprehensive genome annotation

genomatix software suite v2.0

Main menu | Logout

NGS Analysis | Literature & Pathways | Genomes & Data | Pattern Search & Analysis | Pattern Definition | Alignment & Mapping | Tools | Projects & Account | Help



TFBS common to all sequences in a set of six muscle actin promoters

Cartharius et al. *Bioinformatics* 21(13):2933-42

Coming Soon….. HGMD Professional

Coming Soon….. HGMD Professional

Applications of HGMD® include:

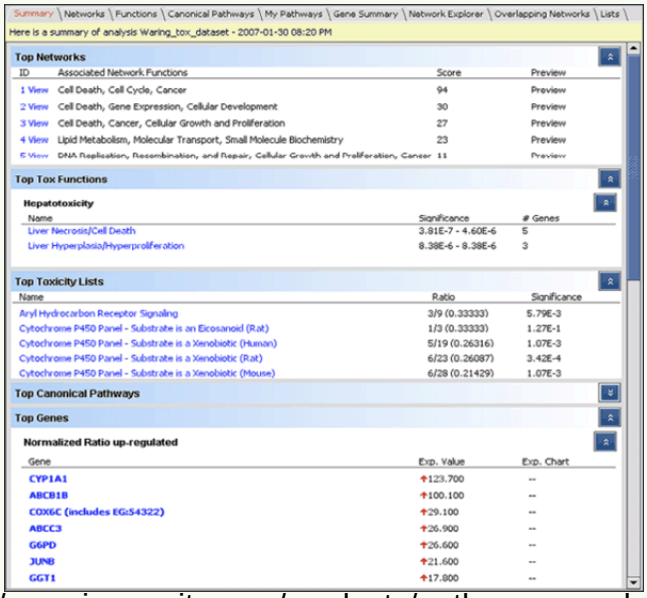
* Determining the novelty of identified gene mutations
* Obtaining all known mutations for a given gene or
  all mutated genes for a disease of interest
* Mapping mutations to a full genome sequence
* Location in a particular motif
  within a splice site or regulatory region

Ingenuity Pathway Analysis

- Search and Explore Biological and Chemical Knowledge
  Genes, drugs, chemicals, protein families, normal cellular and disease processes, and signaling and metabolic pathways
- Dynamic Signaling & Metabolic Pathways
- Analyze and Interpret Data
  such as gene expression, SNP microarrays, proteomics experiments, and gene lists

http://www.ingenuity.com/products/pathways_analysis.html

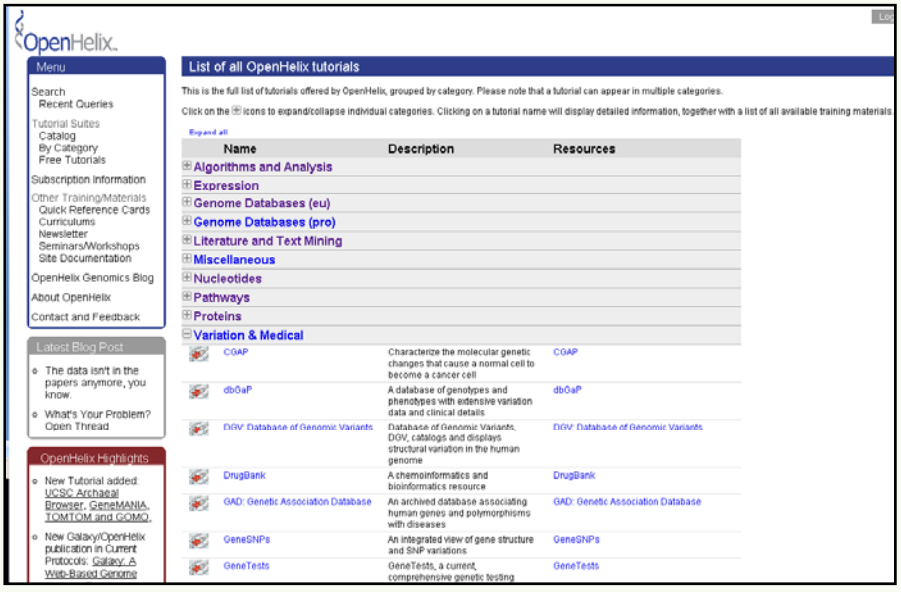
Ingenuity Pathway Analysis



http://www.ingenuity.com/products/pathways_analysis.html

OpenHelix

- Search function to find the right bioinformatics resource
- Tutorial Suites to learn the resource
  - Over 100 on-line self-run narrated tutorials
  - Includes training materials—Powerpoint Slides, Handouts and Exercises
  - Covering a range of needs and research areas
  - Continually updated as resources change
- IP based access at www.openhelix.com



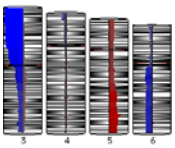
Partek GS™ for Integrated Genomics

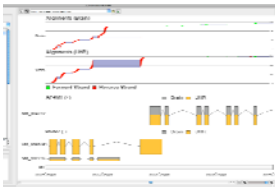
Microarray & Next Generation Sequencing

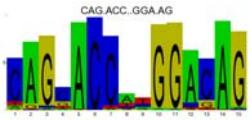
Genome
- Copy Number
  - Total & Allele Specific
- Association
- Loss of Heterozygosity

Transcriptome
- Gene Expression
- Exon/Alternative Splicing
- DGE & mRNA –Seq

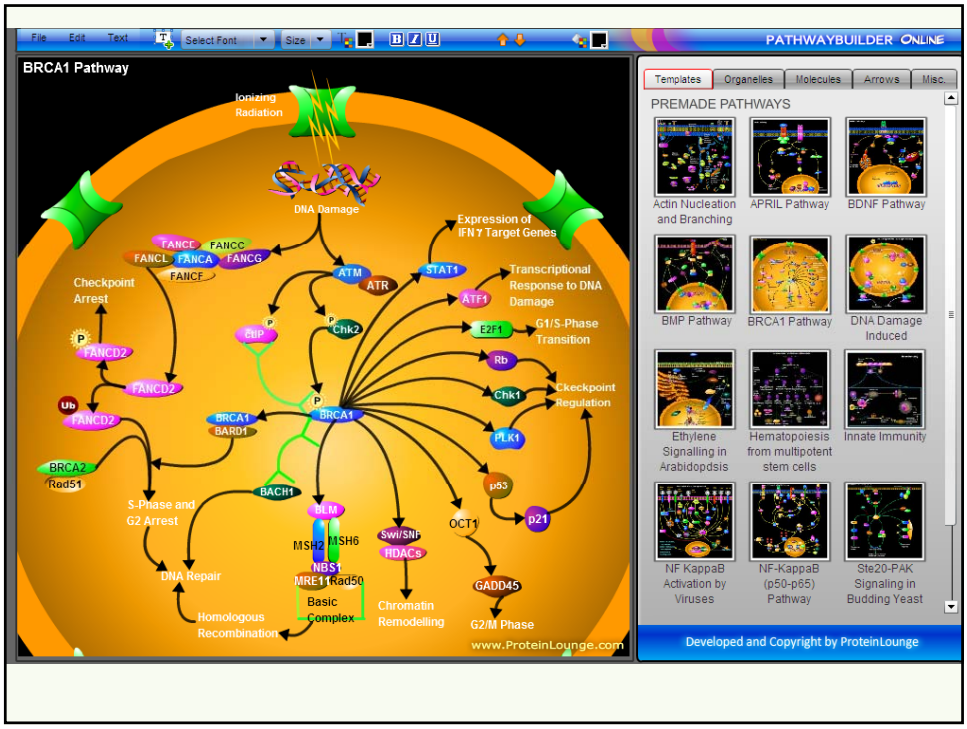
Regulation
- ChIP-Chip
- ChIP-Seq
- microRNA







Partek



Licensed Resources (from the NIH Library)

- DNA/protein sequence analysis
- Expression data analysis
- Pathway analysis
- Next gen sequence analysis
- Promoter/SNP prediction

NIH LIBRARY

Licensed Resources (from the NIH Library)

Floating network seat(s)

to register go to
http://nihlibrary.nih.gov/bioinformatics

Some licensed resources loaded on two dedicated computers in the Library

additional software such as

Cytoscape
Cn3D

Outline

- Databases and tools (free)
- Licensed resources (from the NIH Library)
  Dedicated computers
- Specific examples
- Training and additional help

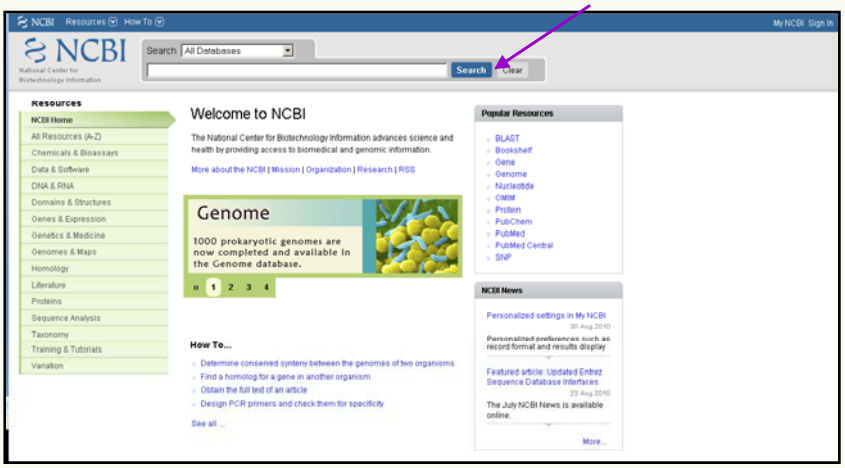
Practical examples

- Obtain human non-synonymous SNPs that are associated with diseases and with known 3-D structures of the proteins
- Identify SNPs in conserved regions
- Download upstream sequences for multiple human genes
- Obtain unique genes in a genome compared to other genome(s)

Practical examples

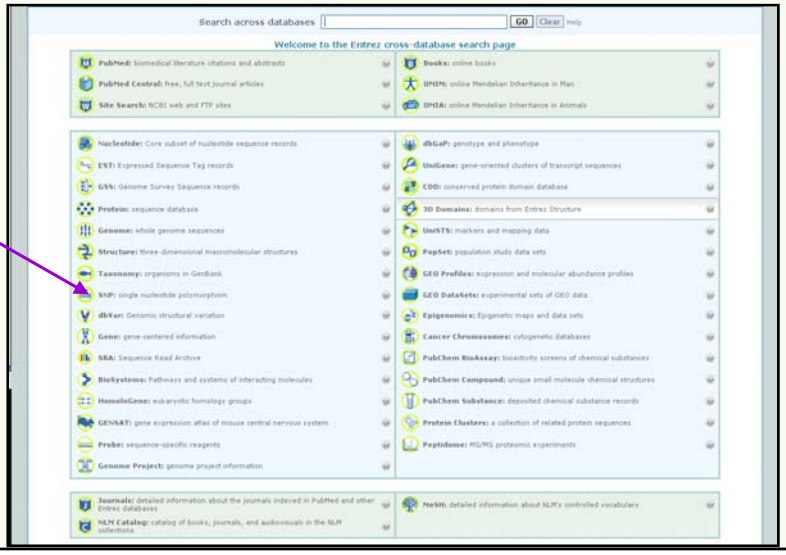
- Non-synonymous SNPs that are associated with diseases and with known 3-D structures of the proteins (NCBI)
- SNPs in the conserved regions (UCSC genome browser)
- Download upstream sequences for multiple human genes (Ensembl)
- Obtain unique genes in a genome compared to other genome(s) (IMG)

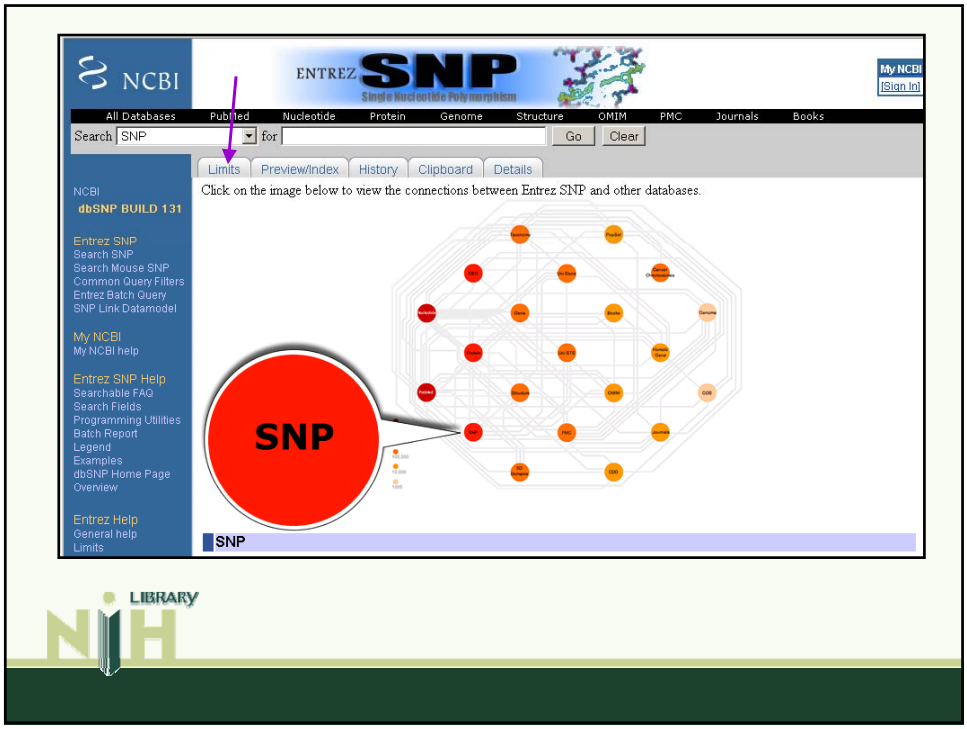
Non-synonymous human SNPs that are associated with diseases and with known 3-D structures of the proteins

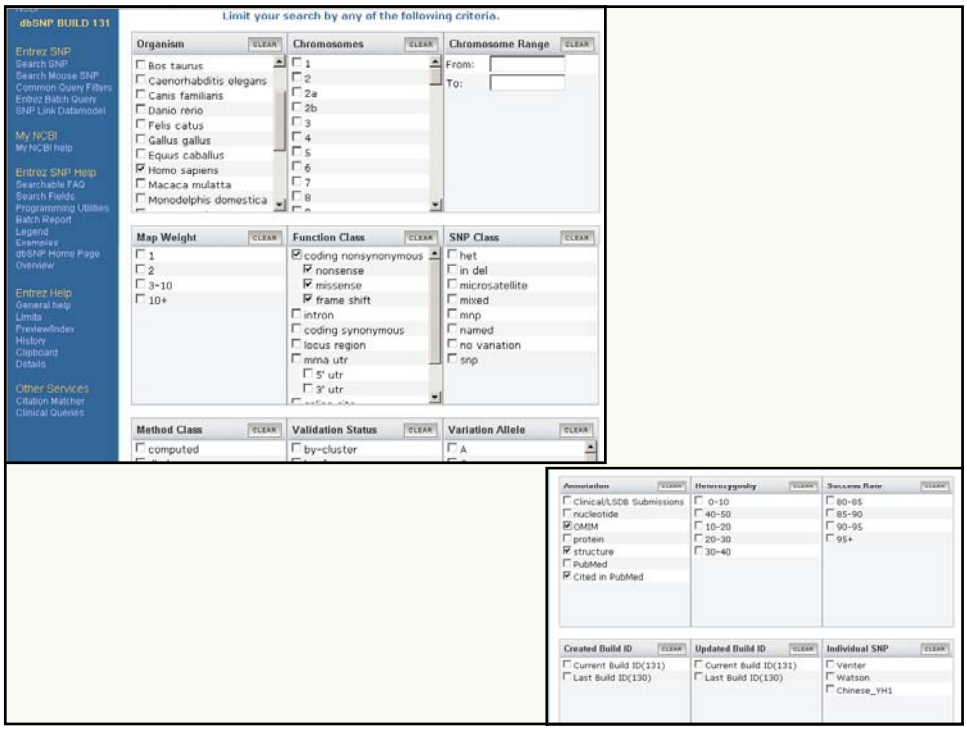


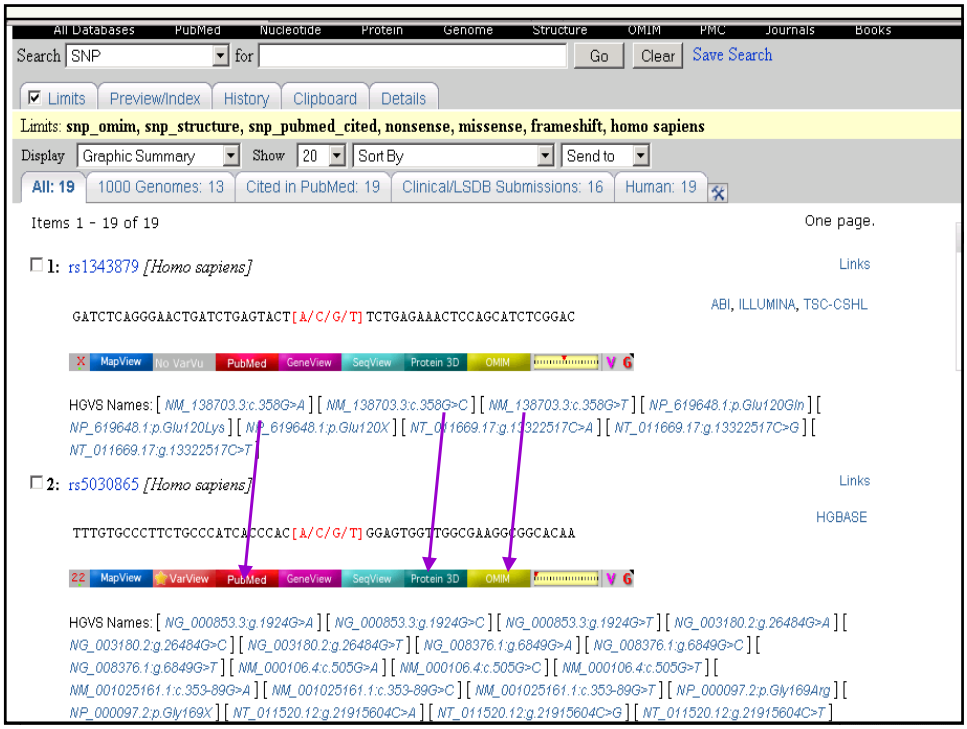
http://www.ncbi.nlm.nih.gov/

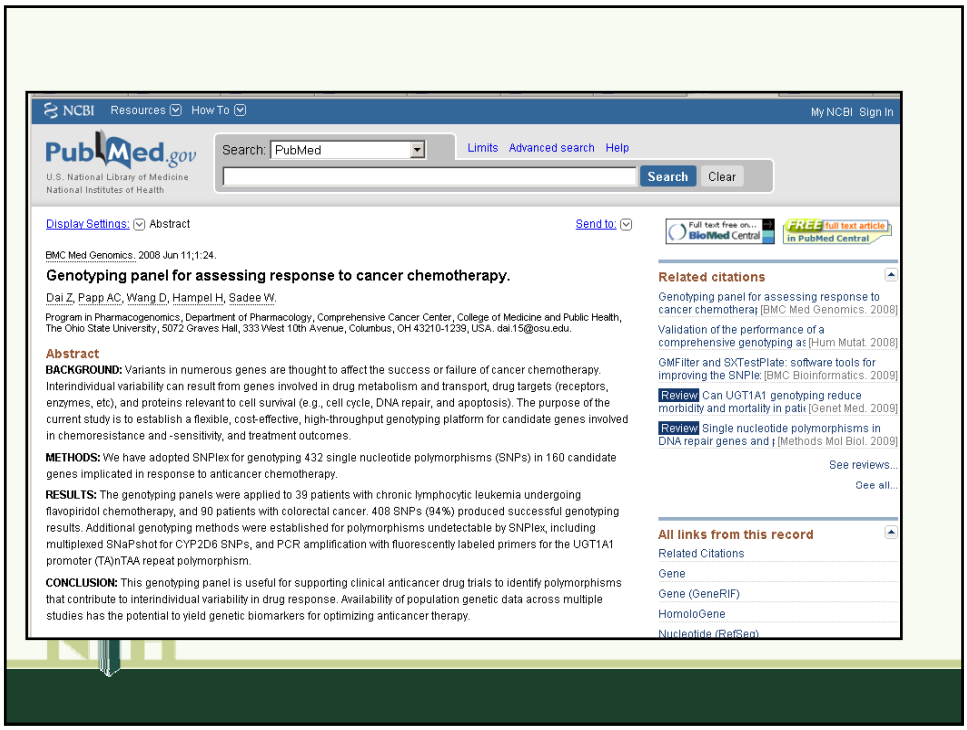
Non-synonymous SNPs that are associated with diseases and with known 3-D structures of the proteins

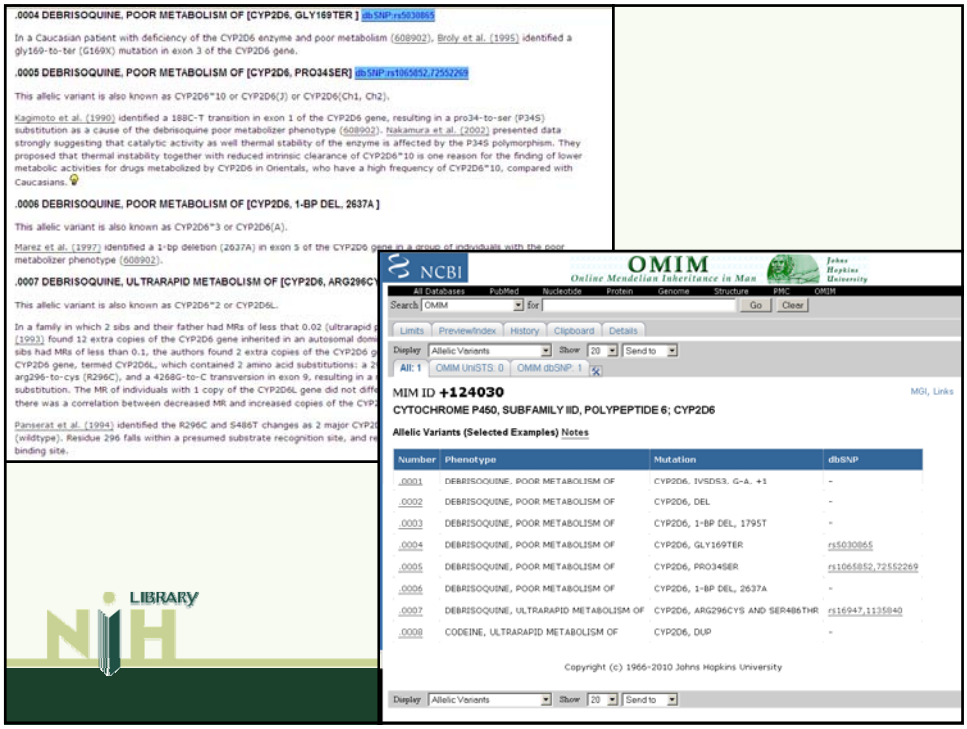


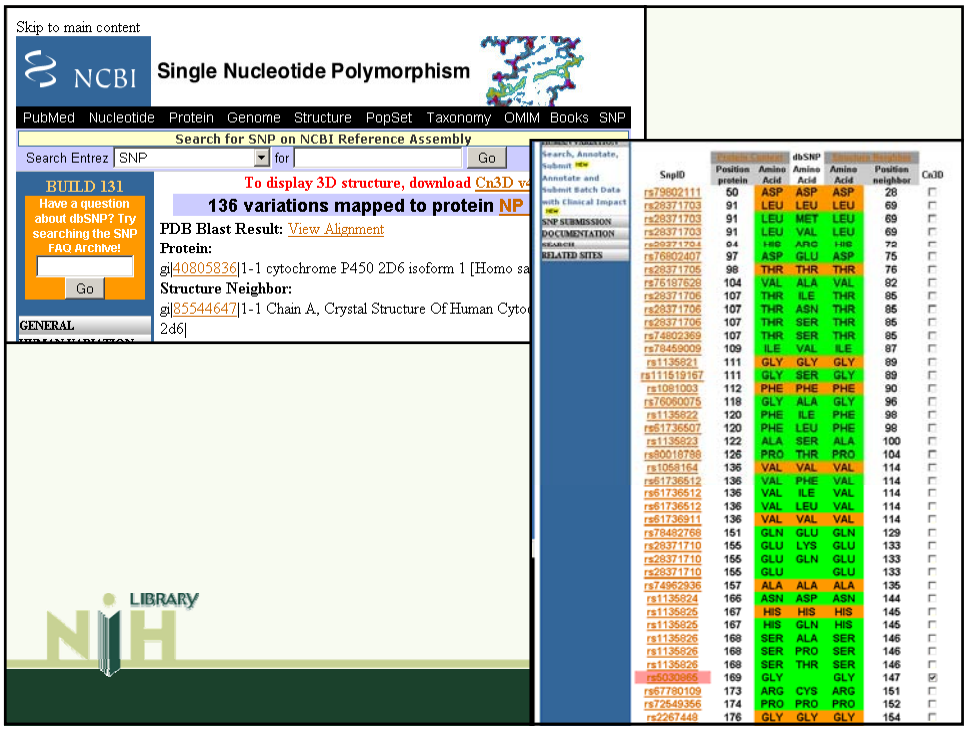

NCBI
ENTREZ SNP
Single Nucleotide Polymorphism
My NCBI
[Sign In]
All Databases
PubMed
Nucleotide
Protein
Genome
Structure
OMIM
PMC
Journals
Books
Search SNP for
Go
Clear
Limits
Preview/Index
History
Clipboard
Details
NCBI
dbSNP BUILD 131
Entrez SNP
Search SNP
Search Mouse SNP
Common Query Filters
Entrez Batch Query
SNP Link Datamodel
My NCBI
My NCBI help
Entrez SNP Help
Searchable FAQ
Search Fields
Programming Utilities
Batch Report
Legend
Examples
dbSNP Home Page
Overview
Entrez Help
General help
Limits
Click on the image below to view the connections between Entrez SNP and other databases.
SNP
SNP
LIBRARY
NIH

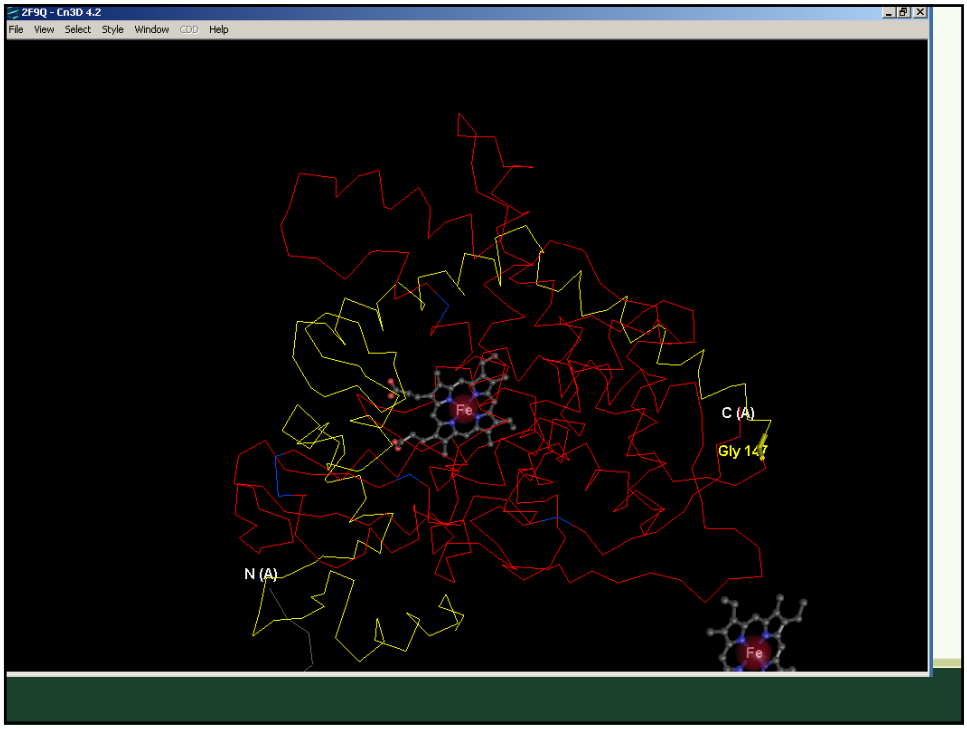

dbSNP BUILD 131
Limit your search by any of the following criteria.
Organism
Chromosomes
Chromosome Range
From:
To:
Map Weight
Function Class
SNP Class
Method Class
Validation Status
Variation Allele
Annotation
Heterozygosity
Success Rate
Created Build ID
Updated Build ID
Individual SNP


All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search SNP for Go Clear Save Search
Limits Preview/Index History Clipboard Details
Limits: snp_omim, snp_structure, snp_pubmed_cited, nonsense, missense, frameshift, homo sapiens
Display Graphic Summary Show 20 Sort By Send to
All: 19 1000 Genomes: 13 Cited in PubMed: 19 Clinical/LSDB Submissions: 16 Human: 19
Items 1 - 19 of 19
One page.
1: rs1343879 [Homo sapiens]
Links
ABI, ILLUMINA, TSC-CSHL
GATCTCAGGGAACTGATCTGAGTACT[A/C/G/T]TCTGAGAAACTCCAGCATCTCGGAC
MapView No VarVu PubMed GeneView SeqView Protein 3D OMIM
HGVS Names: [NM_138703.3:c.358G>A] [NM_138703.3:c.358G>C] [NM_138703.3:c.358G>T] [NP_619648.1:p.Glu120Gln] [NP_619648.1:p.Glu120Lys] [NP_619648.1:p.Glu120X] [NT_011669.17:g.13322517C>A] [NT_011669.17:g.13322517C>G] [NT_011669.17:g.13322517C>T]
2: rs5030865 [Homo sapiens]
Links
HGBASE
TTTGTGCCCTTCTGCCCATCACCCAC[A/C/G/T]GGAGTGGTTGGCGAAGGCGGCACAA
22 MapView VarView PubMed GeneView SeqView Protein 3D OMIM
HGVS Names: [NG_000853.3:g.1924G>A] [NG_000853.3:g.1924G>C] [NG_000853.3:g.1924G>T] [NG_003180.2:g.26484G>A] [NG_003180.2:g.26484G>C] [NG_003180.2:g.26484G>T] [NG_008376.1:g.6849G>A] [NG_008376.1:g.6849G>C] [NG_008376.1:g.6849G>T] [NM_000106.4:c.505G>A] [NM_000106.4:c.505G>C] [NM_000106.4:c.505G>T] [NM_001025161.1:c.353-89G>A] [NM_001025161.1:c.353-89G>C] [NM_001025161.1:c.353-89G>T] [NP_000097.2:p.Gly169Arg] [NP_000097.2:p.Gly169X] [NT_011520.12:g.21915604C>A] [NT_011520.12:g.21915604C>G] [NT_011520.12:g.21915604C>T]

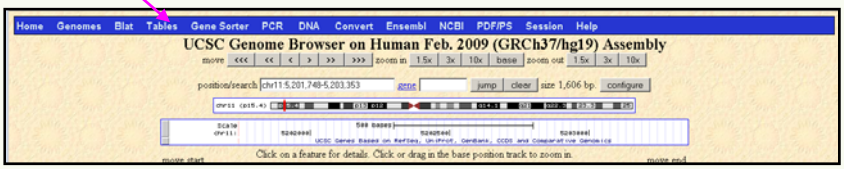

NCBI Resources How To My NCBI Sign In
PubMed.gov
U.S. National Library of Medicine
National Institutes of Health
Search: PubMed Limits Advanced search Help
Search Clear
Display Settings: Abstract
Send to:
BMC Med Genomics. 2008 Jun 11;1:24.
Genotyping panel for assessing response to cancer chemotherapy.
Dai Z, Papp AC, Wang D, Hampel H, Sadee W.
Program in Pharmacogenomics, Department of Pharmacology, Comprehensive Cancer Center, College of Medicine and Public Health, The Ohio State University, 5072 Graves Hall, 333 West 10th Avenue, Columbus, OH 43210-1239, USA. dai.15@osu.edu.
Abstract
BACKGROUND: Variants in numerous genes are thought to affect the success or failure of cancer chemotherapy. Interindividual variability can result from genes involved in drug metabolism and transport, drug targets (receptors, enzymes, etc), and proteins relevant to cell survival (e.g., cell cycle, DNA repair, and apoptosis). The purpose of the current study is to establish a flexible, cost-effective, high-throughput genotyping platform for candidate genes involved in chemoresistance and -sensitivity, and treatment outcomes.
METHODS: We have adopted SNPlex for genotyping 432 single nucleotide polymorphisms (SNPs) in 160 candidate genes implicated in response to anticancer chemotherapy.
RESULTS: The genotyping panels were applied to 39 patients with chronic lymphocytic leukemia undergoing flavopiridol chemotherapy, and 90 patients with colorectal cancer. 408 SNPs (94%) produced successful genotyping results. Additional genotyping methods were established for polymorphisms undetectable by SNPlex, including multiplexed SNaPshot for CYP2D6 SNPs, and PCR amplification with fluorescently labeled primers for the UGT1A1 promoter (TA)nTAA repeat polymorphism.
CONCLUSION: This genotyping panel is useful for supporting clinical anticancer drug trials to identify polymorphisms that contribute to interindividual variability in drug response. Availability of population genetic data across multiple studies has the potential to yield genetic biomarkers for optimizing anticancer therapy.
Related citations
Genotyping panel for assessing response to cancer chemotherap [BMC Med Genomics. 2008]
Validation of the performance of a comprehensive genotyping as [Hum Mutat. 2008]
GMFilter and SXTestPlate: software tools for improving the SNPle [BMC Bioinformatics. 2009]
Review Can UGT1A1 genotyping reduce morbidity and mortality in patie [Genet Med. 2009]
Review Single nucleotide polymorphisms in DNA repair genes and p [Methods Mol Biol. 2009]
See reviews...
See all...
All links from this record
Related Citations
Gene
Gene (GeneRIF)
HomoloGene
Nucleotide (RefSeq)

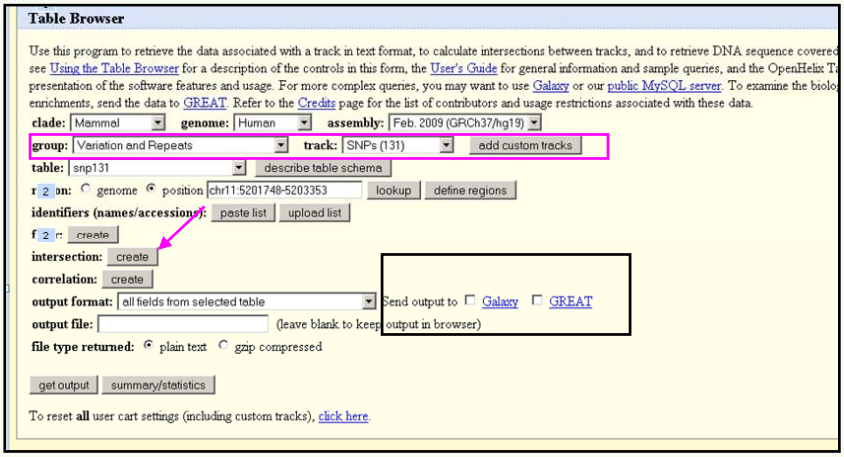

.0004 DEBRISOQUINE, POOR METABOLISM OF [CYP2D6, GLY169TER]
.0005 DEBRISOQUINE, POOR METABOLISM OF [CYP2D6, PRO34SER]
.0006 DEBRISOQUINE, POOR METABOLISM OF [CYP2D6, 1-BP DEL, 2637A]
.0007 DEBRISOQUINE, ULTRARAPID METABOLISM OF [CYP2D6, ARG296CY
NCBI
OMIM
Online Mendelian Inheritance in Man
MIM ID +124030
CYTOCHROME P450, SUBFAMILY IID, POLYPEPTIDE 6; CYP2D6
Allelic Variants (Selected Examples) Notes
Number
Phenotype
Mutation
dbSNP
.0001 DEBRISOQUINE, POOR METABOLISM OF CYP2D6, IVSDS3, G-A, +1
.0002 DEBRISOQUINE, POOR METABOLISM OF CYP2D6, DEL
.0003 DEBRISOQUINE, POOR METABOLISM OF CYP2D6, 1-BP DEL, 1795T
.0004 DEBRISOQUINE, POOR METABOLISM OF CYP2D6, GLY169TER rs5030865
.0005 DEBRISOQUINE, POOR METABOLISM OF CYP2D6, PRO34SER rs1065852,72552269
.0006 DEBRISOQUINE, POOR METABOLISM OF CYP2D6, 1-BP DEL, 2637A
.0007 DEBRISOQUINE, ULTRARAPID METABOLISM OF CYP2D6, ARG296CYS AND SER486THR rs16947,1135840
.0008 CODEINE, ULTRARAPID METABOLISM OF CYP2D6, DUP
Copyright (c) 1966-2010 Johns Hopkins University
LIBRARY
NIH

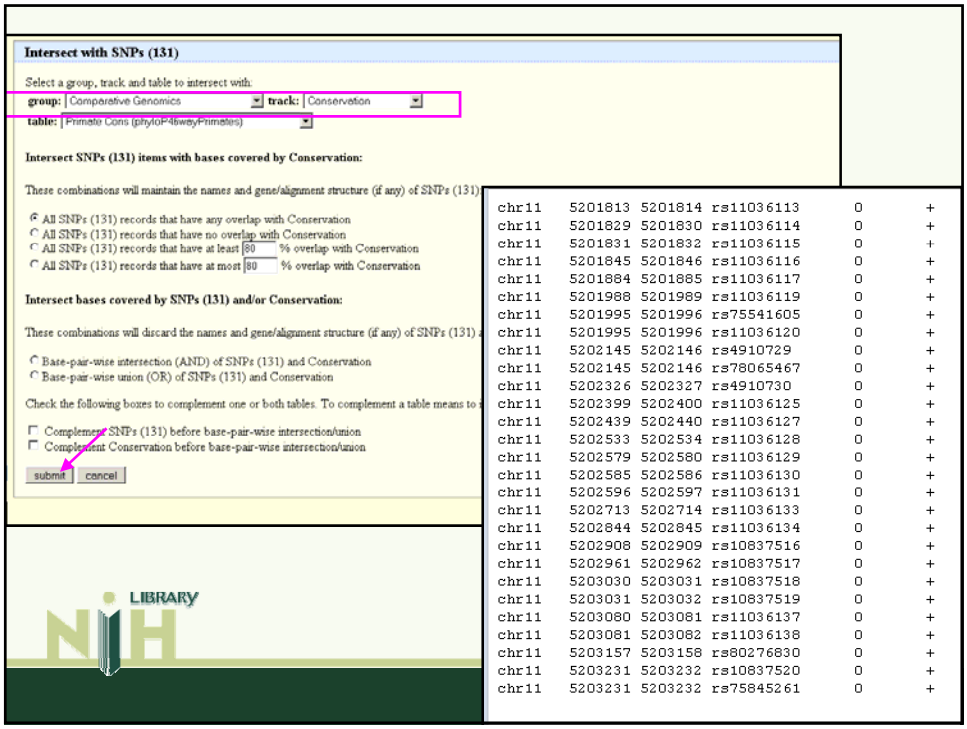

Skip to main content
NCBI
Single Nucleotide Polymorphism
PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP
Search for SNP on NCBI Reference Assembly
Search Entrez SNP for Go
BUILD 131
Have a question about dbSNP? Try searching the SNP FAQ Archive!
Go
GENERAL
To display 3D structure, download Cn3D v
136 variations mapped to protein NP
PDB Blast Result: View Alignment
Protein:
gi|40805836|1-1 cytochrome P450 2D6 isoform 1 [Homo sa
Structure Neighbor:
gi|85544647|1-1 Chain A, Crystal Structure Of Human Cyto
2d6|
Search, Annotate, Submit NEW
Annotate and Submit Batch Data with Clinical Impact NEW
SNP SUBMISSION
DOCUMENTATION
SEARCH
RELATED SITES
Protein Context
dbSNP
Structure Neighbor
SnpID
Position protein
Amino Acid
Amino Acid
Amino Acid
Position neighbor
Cn3D
rs79802111 50 ASP ASP ASP 28
rs28371703 91 LEU LEU LEU 69
rs28371703 91 LEU MET LEU 69
rs28371703 91 LEU VAL LEU 69
rs74802407 97 ASP GLU ASP 75
rs28371705 98 THR THR THR 76
rs76187628 104 VAL ALA VAL 82
rs28371706 107 THR ILE THR 85
rs28371706 107 THR ASN THR 85
rs28371706 107 THR SER THR 85
rs74802369 107 THR SER THR 85
rs78459009 109 ILE VAL ILE 87
rs1135821 111 GLY GLY GLY 89
rs111519167 111 GLY SER GLY 89
rs1081003 112 PHE PHE PHE 90
rs76060075 118 GLY ALA GLY 96
rs1135822 120 PHE ILE PHE 98
rs61736507 120 PHE LEU PHE 98
rs1135823 122 ALA SER ALA 100
rs80016788 126 PRO THR PRO 104
rs1058164 136 VAL VAL VAL 114
rs61736512 136 VAL PHE VAL 114
rs61736512 136 VAL ILE VAL 114
rs61736512 136 VAL LEU VAL 114
rs61736911 136 VAL VAL VAL 114
rs78482768 151 GLN GLU GLN 129
rs28371710 155 GLU LYS GLU 133
rs28371710 155 GLU GLN GLU 133
rs28371710 155 GLU GLU 133
rs74962936 157 ALA ALA ALA 135
rs1135824 166 ASN ASP ASN 144
rs1135825 167 HIS HIS HIS 145
rs1135825 167 HIS GLN HIS 145
rs1135826 168 SER ALA SER 146
rs1135826 168 SER PRO SER 146
rs1135826 168 SER THR SER 146
rs5030865 169 GLY GLY 147
rs67780109 173 ARG CYS ARG 151
rs72549356 174 PRO PRO PRO 152
rs2267448 176 GLY GLY GLY 154
LIBRARY
NIH



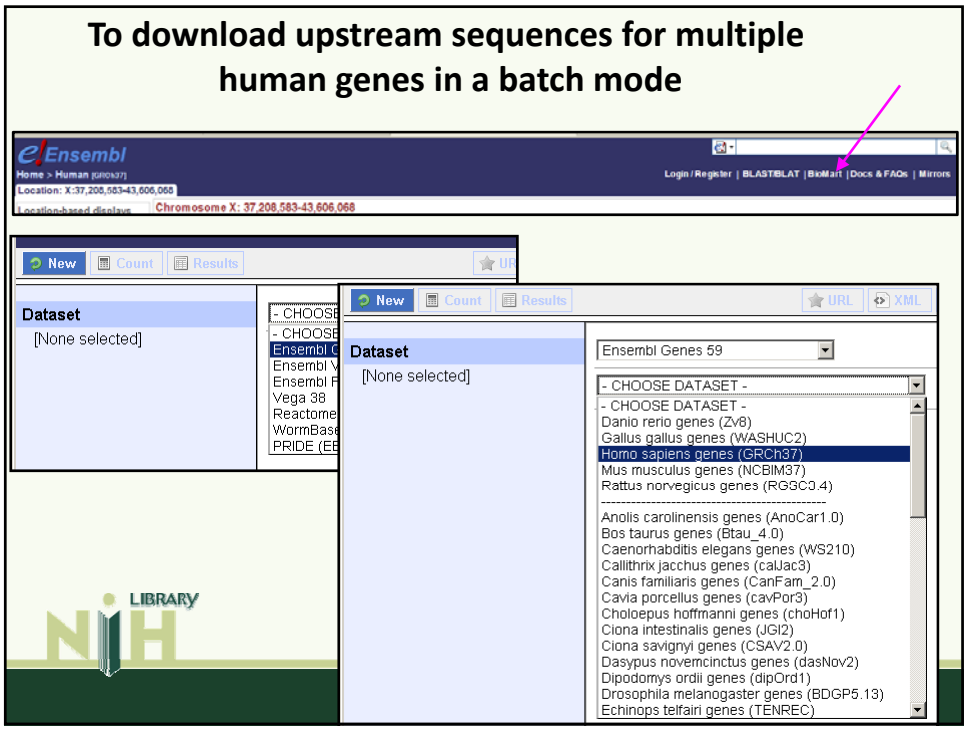
To download SNPs in conserved regions



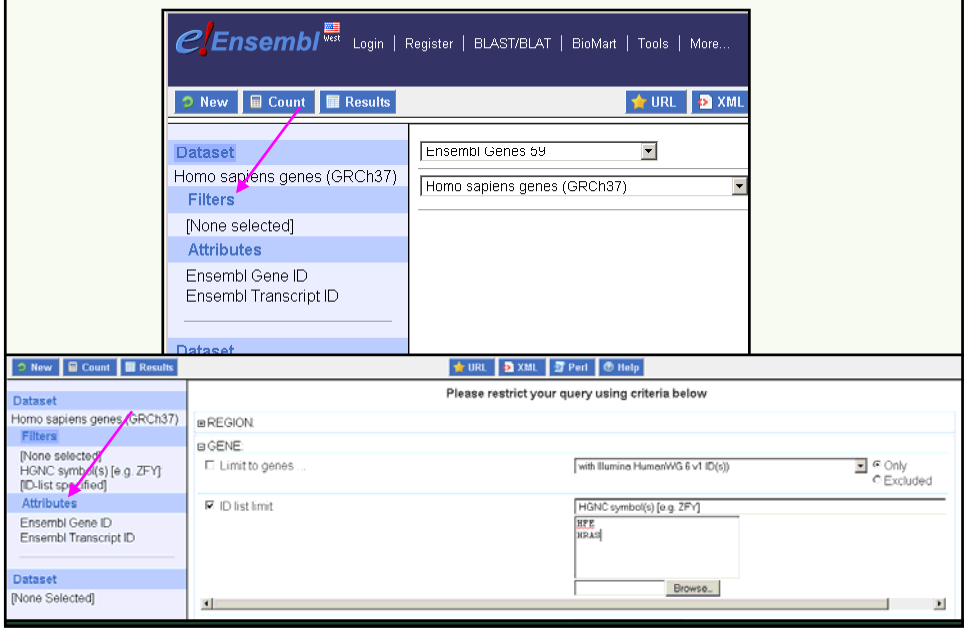


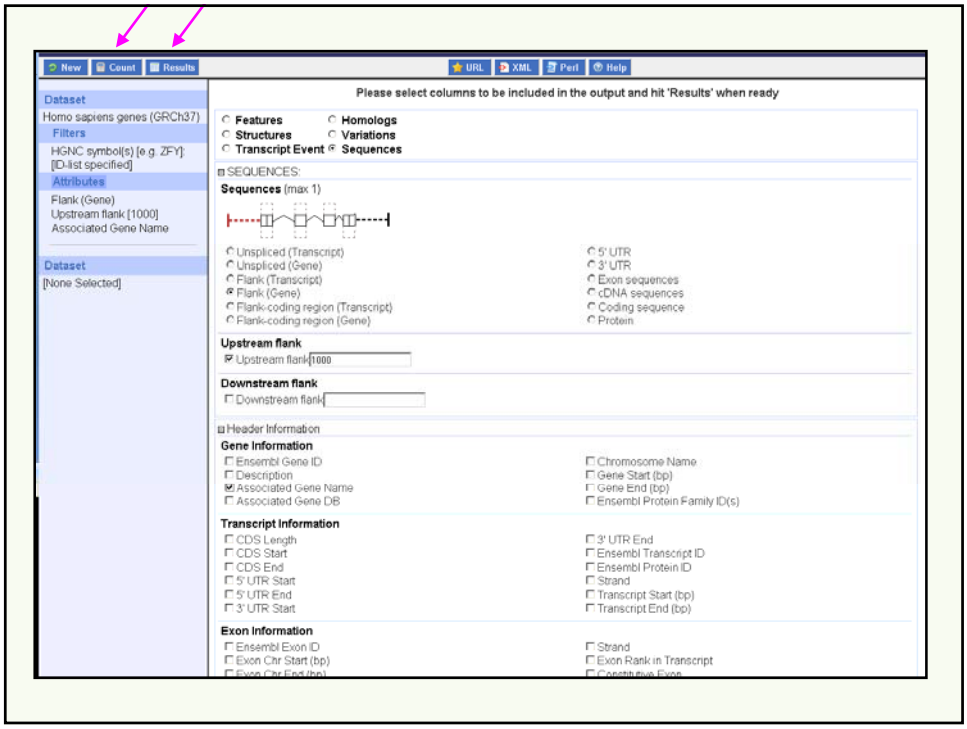


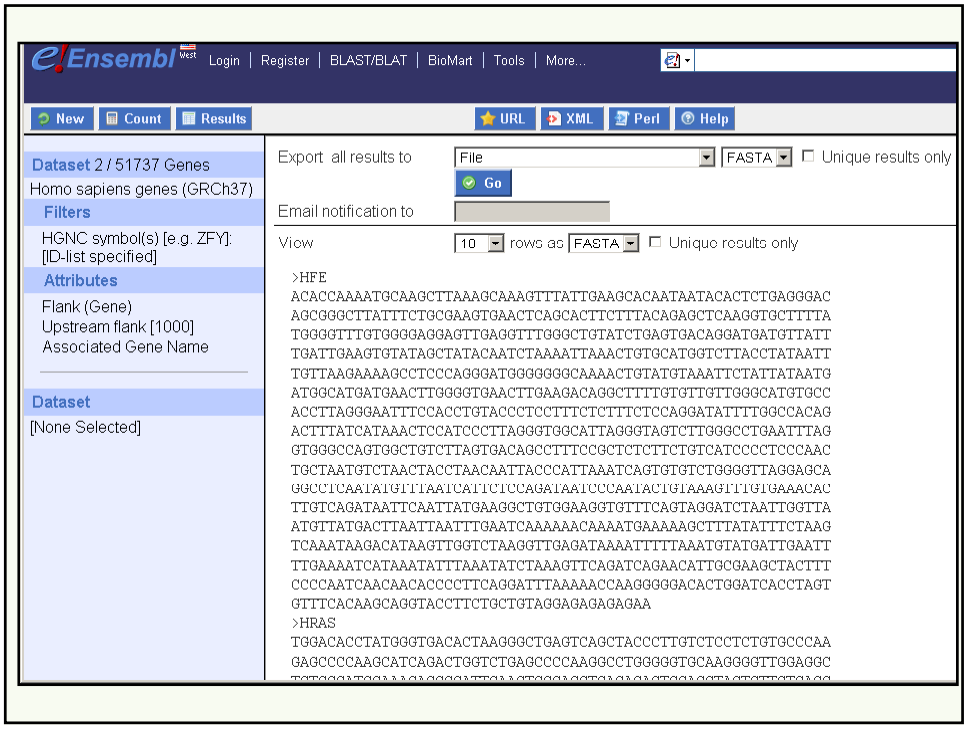
To download upstream sequences for multiple human genes in a batch mode

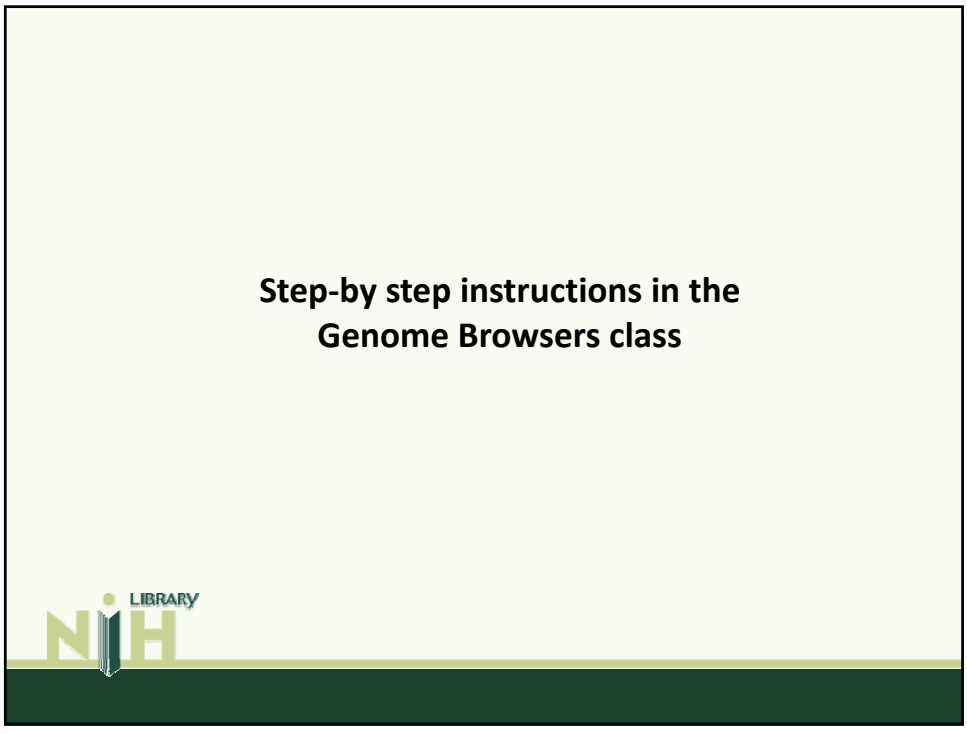


To download upstream sequences for multiple human genes in a batch mode

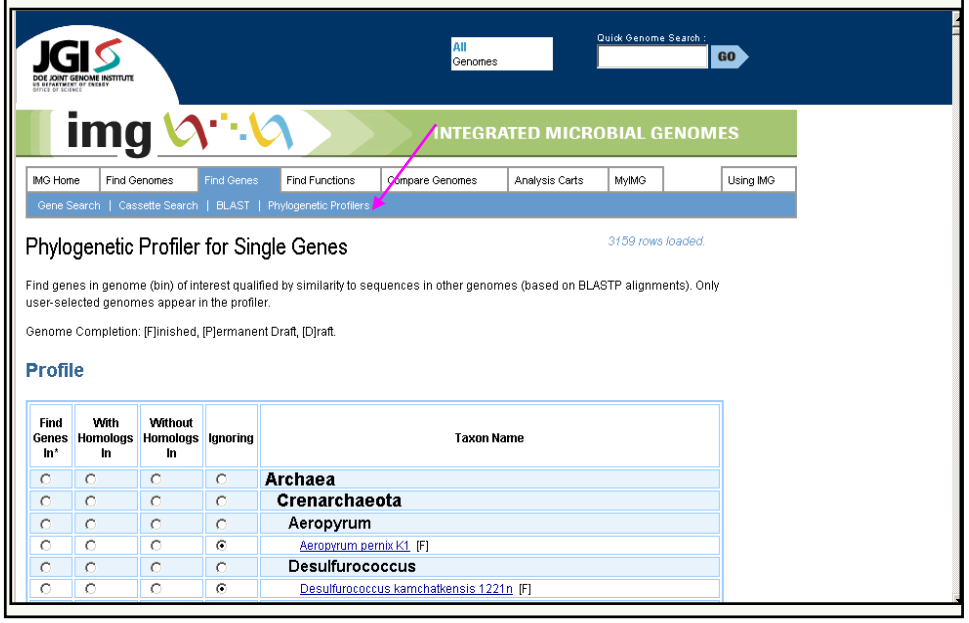


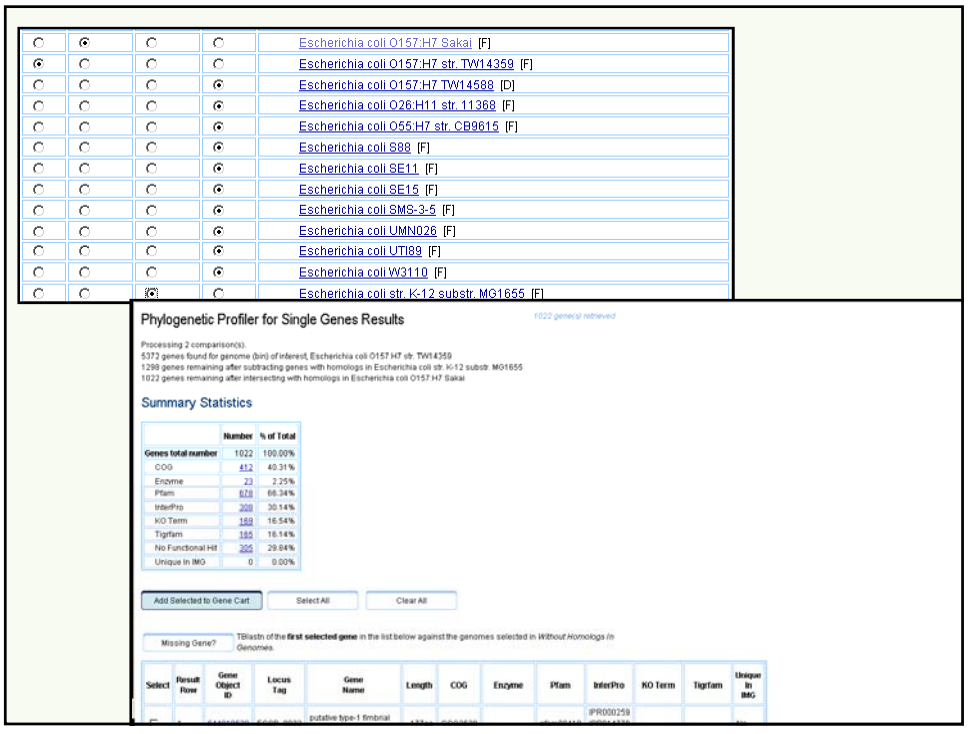


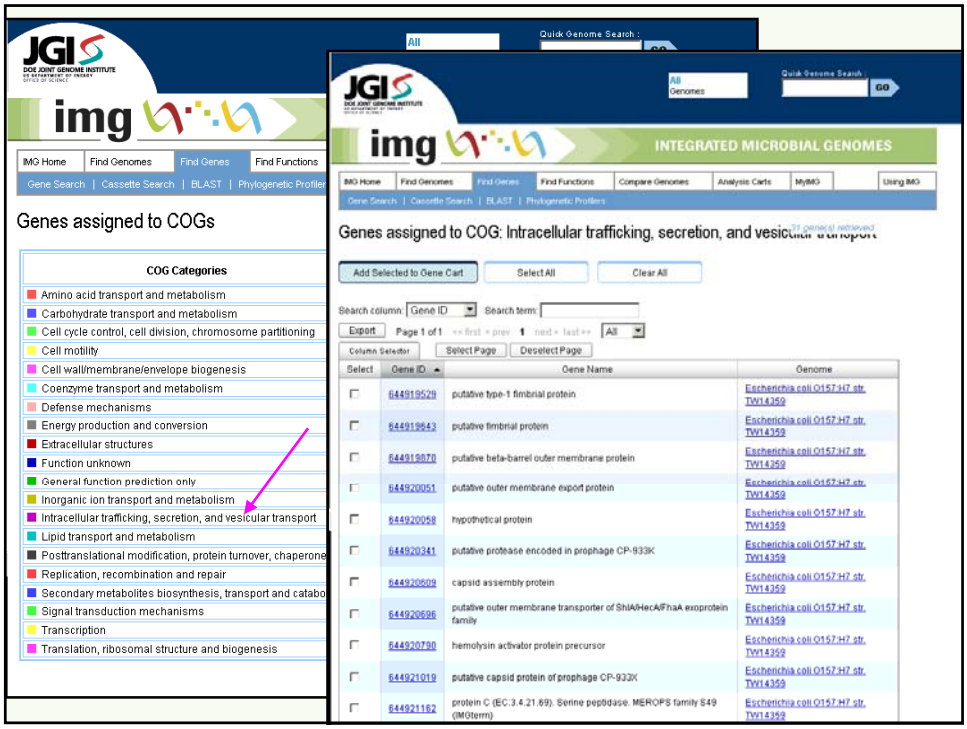

Ensembl
Login | Register | BLAST/BLAT | BioMart | Tools | More...
New
Count
Results
URL
XML
Perl
Help
Dataset 2 / 51737 Genes
Homo sapiens genes (GRCh37)
Filters
HGNC symbol(s) [e.g. ZFY]: [ID-list specified]
Attributes
Flank (Gene)
Upstream flank [1000]
Associated Gene Name
Dataset
[None Selected]
Export all results to
File
FASTA
Unique results only
Go
Email notification to
View
10
rows as
FASTA
Unique results only
>HFE
ACACCAAAATGCAAGCTTAAAGCAAAGTTTATTGAAGCACAATAATACACTCTGAGGGAC
AGCGGGCTTATTTCTGCGAAGTGAACTCAGCACTTCTTTACAGAGCTCAAGGTGCTTTTA
TGGGGTTTGTGGGGAGGAGTTGAGGTTTGGGCTGTATCTGAGTGACAGGATGATGTTATT
TGATTGAAGTGTATAGCTATACAATCTAAAATTAAACTGTGCATGGTCTTACCTATAATT
TGTTAAGAAAAGCCTCCCAGGGATGGGGGGGCAAAACTGTATGTAAATTCTATTATAATG
ATGGCATGATGAACTTGGGGTGAACTTGAAGACAGGCTTTTGTGTTGTTGGGCATGTGCC
ACCTTAGGGAATTTCCACCTGTACCCTCCTTTCTCTTTCTCCAGGATATTTTGGCCACAG
ACTTTATCATAAACTCCATCCCTTAGGGTGGCATTAGGGTAGTCTTGGGCCTGAATTTAG
GTGGGCCAGTGGCTGTCTTAGTGACAGCCTTTCCGCTCTCTTCTGTCATCCCCTCCCAAC
TGCTAATGTCTAACTACCTAACAATTACCCATTAAATCAGTGTGTCTGGGGTTAGGAGCA
GGCCTCAATATGTTAATCATTCTCCAGATAATCCCAATACTGTAAAGTTTGTGAAACAC
TTGTCAGATAATTCAATTATGAAGGCTGTGGAAGGTGTTTCAGTAGGATCTAATTGGTTA
ATGTTATGACTTAATTAATTTGAATCAAAAAACAAAATGAAAAAGCTTTATATTTCTAAG
TCAAATAAGACATAAGTTGGTCTAAGGTTGAGATAAAATTTTTAAATGTATGATTGAATT
TTGAAAATCATAAATATTTAAATATCTAAAGTTCAGATCAGAACATTGCGAAGCTACTTT
CCCCAATCAACAACACCCCTTCAGGATTTAAAAACCAAGGGGGACACTGGATCACCTAGT
GTTTCACAAGCAGGTACCTTCTGCTGTAGGAGAGAGAGAA
>HRAS
TGGACACCTATGGGTGACACTAAGGGCTGAGTCAGCTACCCTTGTCTCCTCTGTGCCCAA
GAGCCCCAAGCATCAGACTGGTCTGAGCCCCAAGGCCTGGGGGTGCAAGGGGTTGGAGGC


Step-by step instructions in the Genome Browsers class
LIBRARY
NIH

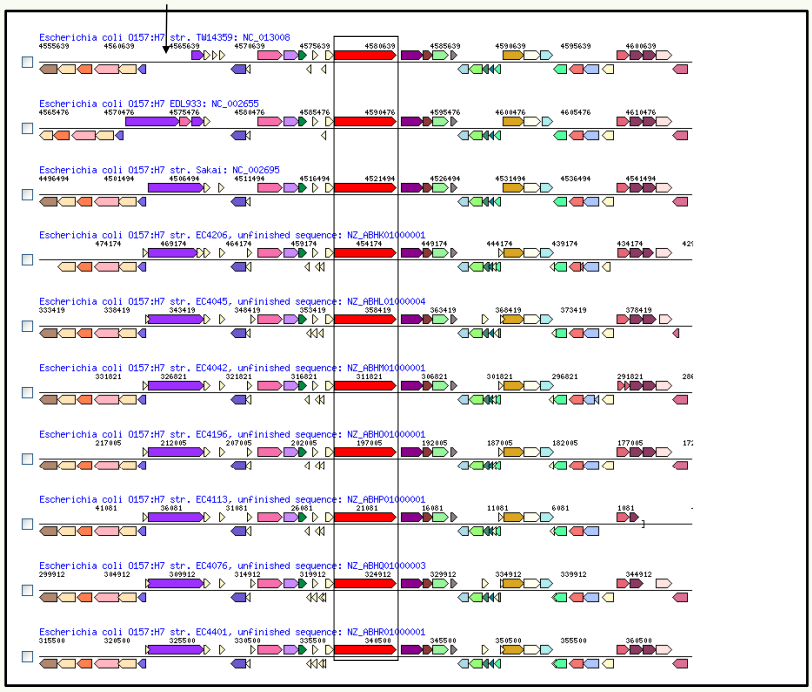
Obtain unique genes in a genome compared to other genome(s)






JGI
DOE JOINT GENOME INSTITUTE
img
IMG Home
Find Genomes
Find Genes
Find Functions
Gene Search | Cassette Search | BLAST | Phylogenetic Profiler
Genes assigned to COGs
COG Categories
Amino acid transport and metabolism
Carbohydrate transport and metabolism
Cell cycle control, cell division, chromosome partitioning
Cell motility
Cell wall/membrane/envelope biogenesis
Coenzyme transport and metabolism
Defense mechanisms
Energy production and conversion
Extracellular structures
Function unknown
General function prediction only
Inorganic ion transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Lipid transport and metabolism
Posttranslational modification, protein turnover, chaperone
Replication, recombination and repair
Secondary metabolites biosynthesis, transport and catabo
Signal transduction mechanisms
Transcription
Translation, ribosomal structure and biogenesis
Quick Genome Search
INTEGRATED MICROBIAL GENOMES
Genes assigned to COG: Intracellular trafficking, secretion, and vesicu
Add Selected to Gene Cart
Select All
Clear All
Export
Select Page
Deselect Page
Gene ID
Gene Name
Genome
putative type-1 fimbrial protein
putative fimbrial protein
putative beta-barrel outer membrane protein
putative outer membrane export protein
hypothetical protein
putative protease encoded in prophage CP-933K
capsid assembly protein
putative outer membrane transporter of ShlA/HecA/FhaA exoprotein family
hemolysin activator protein precursor
putative capsid protein of prophage CP-933X
protein C (EC:3.4.21.69); Serine peptidase. MEROPS family S49 (IMGterm)
Escherichia coli O157:H7 str. TW14359


Evidence For Function Prediction
Neighborhood
red = Current Gene
green = Positional Cluster Gene in the same KEGG Pathway as the Current Gene
CRISPR array
Sequence Viewer For Alternate ORF Search
Chromosome Viewer colored by
— Select Function —
Conserved Neighborhood
Show ortholog neighborhood regions
Chromosomal Cassette Viewer By
– Select Protein Cluster –
COG
COG ID
Consensus Sequence Length
Description
Percent Identity
Alignment On Query Gene
E-value
Bit Score
COG5295
716
[M] Extracellular structures
[U] Intracellular trafficking, secretion, and vesicular transport
Autotransporter adhesin
34.8
2.0e-35
146
Pfam
Pfam Domain
HMM Pfam Hit
Description
Percent Alignment On Query Gene
Alignment On Query Gene
E-value
HMM Score
Hep_Hag
pfam05658
1.76
2.7e-04
20
1.4e-05
24
2.3e-05
24
1.7e-05
24
1.2e-06
28
4.0e-06
26
9.8e-06
25
5.2e-05
22
HIM
pfam05662
Haemagglutinin
1.51
4.1e-06
26
1.7e-07
30
8.2e-08
31
7.2e-08
31
9.0e-07
28



Practical examples

- Non-synonymous SNPs that are associated with diseases and with known 3-D structures of the proteins (NCBI)
- SNPs in the conserved regions (UCSC)
- Download upstream sequences for multiple human genes (Ensembl)
- Obtain unique genes in a genome compared to other genome(s) (IMG)

NIH LIBRARY

Training

Making Sense of DNA and Protein Sequences

Gene Resources: From Transcription Factor Binding Sites to Function

Sequence Similarity Search: BLAST

Protein Structural Analysis: Binding Sites to Distant Homologs

Genome Browsers

Identification of Disease Genes

Correlation of Disease Genes to Phenotypes

Currently offering as an FAES course

"Practical Bioinformatics"

Training

September 30	GeneSpring 11.0.2
October 27-28	Partek Genomics Suite
November 2	Lasergene and SeqMan Ngen
	ArrayStar/Qseq (DNASTAR)
November 9-10	Genomatix

http://nihlibrary.nih.gov/bioinformatics

One-on-one consultations

Wide range of question

Identifying a correct resource

Collaboration on a research project

Gene set enrichment analysis from microarray experiments
Download upstream gene sequence and identify transcription factor binding sites

Bioinformatics specialist with programming expertise
Dr. Lynn Young

Microarray data analysis
Next gen sequence analysis

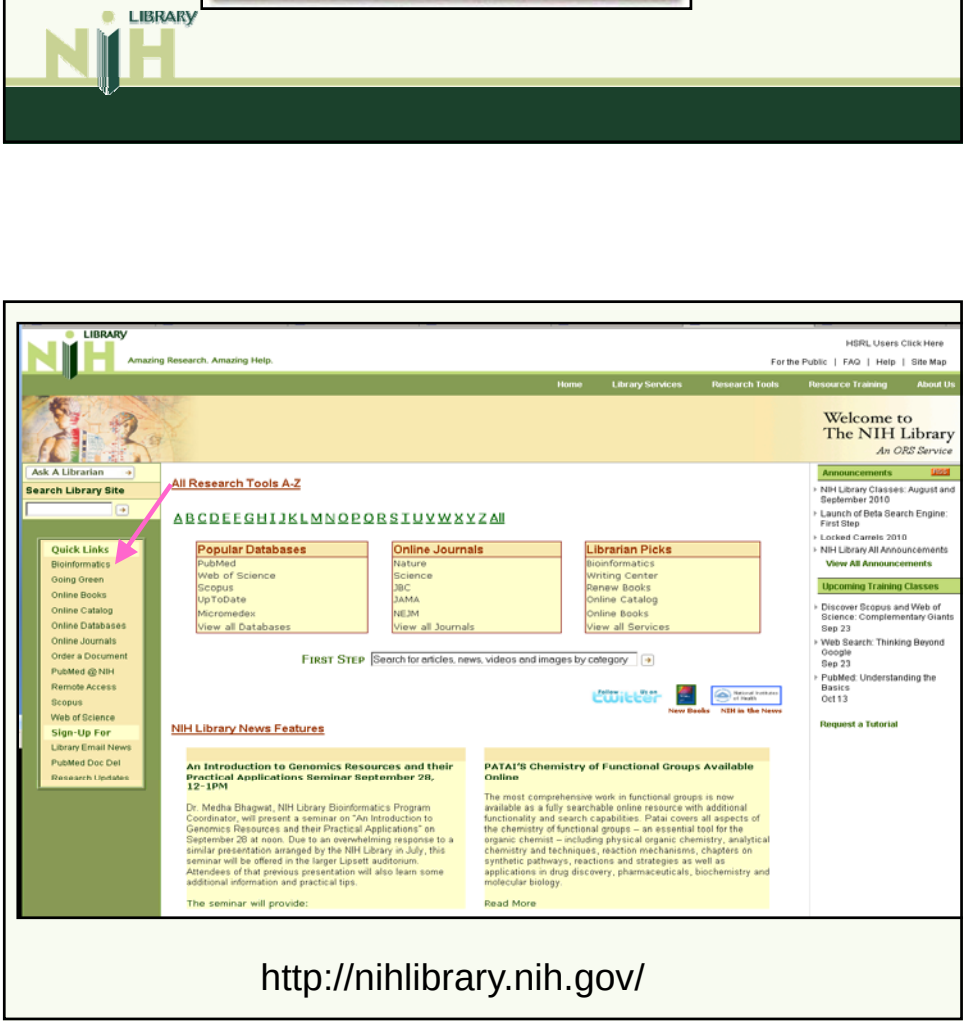
Outline

- Databases and Tools
- Licensed resources from the NIH Library
  - Dedicated computers
- Practical examples
- Training
- One-on-one consults
  - Programming help

Suggestions/Questions/Comments

Medha Bhagwat
bhagwat@mail.nih.gov

Opening New Doors to Information



http://nihlibrary.nih.gov/

http://nihlibrary.nih.gov/bioinformatics

Thanks!

Medha Bhagwat
bhagwat@mail.nih.gov